

# **UNIVERSIDAD NACIONAL DE INGENIERÍA**

**FACULTAD DE INGENIERÍA INDUSTRIAL Y DE SISTEMAS**

**UNIDAD DE POSGRADO**



## **TESIS**

**“MODELO DE INTELIGENCIA DE NEGOCIOS BASADO  
EN MINERÍA DE OPINIÓN SOBRE REDES SOCIALES  
PARA LA MEJORA DEL SERVICIO AL CLIENTE EN UNA  
EMPRESA CORREDORA DE SEGUROS EN EL PERÚ”**

**PARA OBTENER EL GRADO ACADÉMICO DE  
MAESTRO EN CIENCIAS EN  
INGENIERÍA DE SISTEMAS**

**ELABORADO POR:  
ING. SANDRO ALEX CHAUCA PAREDES**

**ASESOR:  
MG. JAVIER TOLENTINO CANCHANO CARO**

**LIMA - PERU**

**2023**

## **DEDICATORIA**

*A mis padres, por apoyarme a recorrer el camino hasta donde me encuentro.*

*A mi esposa, por su motivación y compañía quien siempre me incentivó a  
alcanzar mis objetivos.*

## **AGRADECIMIENTOS**

*A mis asesores de tesis, por todo el apoyo brindado durante el desarrollo de la investigación.*

*A todos mis profesores de la UNI, por todo el conocimiento y experiencia impartida en las aulas que me ha permitido mejorar como profesional.*

## ÍNDICE DE CONTENIDOS

DEDICATORIA .....	i
AGRADECIMIENTOS.....	ii
ÍNDICE DE CONTENIDOS .....	iii
ÍNDICE DE TABLAS .....	xvi
ÍNDICE DE FIGURAS.....	xviii
RESUMEN.....	xxv
ABSTRACT.....	xxx
INTRODUCCIÓN.....	1
CAPÍTULO I.....	3
PROBLEMA DE INVESTIGACIÓN.....	3
1.1. ANTECEDENTES DE LA INVESTIGACIÓN .....	3
1.1.1. Toma de decisiones .....	3
1.1.2. Inteligencia de Negocios .....	4
1.1.3. Redes sociales.....	5
1.1.4. Innovación de productos .....	9

1.1.5.	Big Data .....	11
1.1.6.	Índice de promotores neto (NPS) .....	13
1.2.	DESCRIPCIÓN DE LA REALIDAD PROBLEMÁTICA.....	16
1.2.1.	Inversión en publicidad digital .....	22
1.2.2.	Tenencia de seguro de salud .....	23
1.2.3.	Población ocupada con seguros de salud.....	27
1.2.4.	Incremento del empleo informal .....	30
1.2.5.	Baja profundidad de la inclusión financiera .....	30
1.2.6.	El futuro del seguro de automóvil conectado .....	31
1.3.	FORMULACIÓN DEL PROBLEMA .....	32
1.3.1.	Problema general.....	34
1.3.2.	Problema específico.....	34
1.3.3.	Diagrama de Ishikawa.....	36
1.4.	DELIMITACIÓN DE LA INVESTIGACIÓN.....	37
1.5.	JUSTIFICACIÓN E IMPORTANCIA DE LA INVESTIGACIÓN .....	37
1.6.	OBJETIVOS DE LA INVESTIGACIÓN .....	38
1.6.1.	Objetivo general .....	38
1.6.2.	Objetivos específicos .....	38
1.7.	HIPÓTESIS DE LA INVESTIGACIÓN .....	39
1.7.1.	Hipótesis general.....	39

1.7.2. Hipótesis específicas.....	39
1.8. VARIABLES DE INDICADORES.....	40
1.8.1. Identificación de las variables .....	40
1.8.2. Operacionalización de variables.....	41
1.8.3. Matriz de consistencia.....	42
CAPÍTULO II.....	44
MARCO DE REFERENCIA.....	44
2.1. ANTECEDENTES DE LA INVESTIGACIÓN .....	44
2.1.1. El sector asegurador en el Perú.....	53
2.1.1.1. Análisis del sector asegurador.....	53
2.1.1.1.1. Empresas de seguros.....	53
2.1.1.1.2. Empresas de corredores de seguros.....	59
2.1.1.2. Perfil del consumidor de seguros.....	60
2.2. MARCO TEÓRICO.....	62
2.2.1. Sistemas de información .....	62
2.2.1.1. Dimensiones de los sistemas de información.....	63
2.2.1.2. Tipos de sistemas de información .....	64
2.2.1.2.1. Sistemas de procesamiento de transacciones .....	65
2.2.1.2.2. Sistemas de Inteligencia de Negocios.....	66
2.2.2. Sistemas de Inteligencia de Negocios (BI).....	68

2.2.2.1.	Arquitectura Inteligencia de Negocios (BI).....	69
2.2.2.1.1.	Sistema de soporte operaciones de negocio.....	70
2.2.2.1.2.	Extracción, transformación y carga (ETL) .....	71
2.2.2.1.3.	Área de presentación .....	71
2.2.2.1.4.	Aplicaciones de Inteligencia de Negocios (BI).....	71
2.2.2.2.	Modelamiento dimensional .....	72
2.2.2.2.1.	Tabla de hechos .....	74
2.2.2.2.2.	Tabla de dimensiones .....	75
2.2.2.2.3.	Claves sustitutas de dimensión .....	76
2.2.2.2.4.	Esquema estrella.....	77
2.2.2.2.5.	Diseño dimensional .....	79
2.2.3.	Inteligencia de Negocios Social (SBI) .....	80
2.2.3.1.	Arquitectura Inteligencia de Negocios Social.....	81
2.2.4.	Machine Learning.....	83
2.2.4.1.	Tipos de Machine Learning.....	85
2.2.4.1.1.	Aprendizaje supervisado .....	85
2.2.4.1.1.1.	Support Vector Machine (SVM).....	86
2.2.4.1.2.	Aprendizaje no supervisado .....	89
2.2.5.	Minería de texto.....	91
2.2.5.1.	Minería de texto y minería de datos.....	91

2.2.5.2.	Clasificación de texto.....	93
2.2.5.2.1.	Indexación de texto .....	96
2.2.5.2.1.1.	Tokenización .....	97
2.2.5.2.1.2.	Stemming .....	98
2.2.5.2.1.3.	Stop-word removal .....	99
2.2.5.2.1.4.	Term weighting.....	100
2.2.5.3.	Aplicaciones de clasificación de texto.....	101
2.2.5.4.	Enfoque basado en reglas .....	102
2.2.5.5.	Enfoque basado en Machine Learning .....	103
2.2.5.5.1.	Conjunto de entrenamiento .....	105
2.2.5.6.	Análisis de sentimiento .....	106
2.2.5.7.	Minería de texto en redes sociales .....	111
2.2.6.	Índice de promotores neto (NPS).....	114
2.2.7.	Interfaz de programación de aplicaciones (API).....	116
2.2.8.	RESTful API .....	117
2.2.9.	API Graph Facebook.....	119
2.2.9.1.	Estructura de la API Graph.....	119
2.3.	MARCO CONCEPTUAL.....	120
2.4.	MARCO LEGAL.....	122
CAPÍTULO III	.....	124

METODOLOGÍA DE LA INVESTIGACIÓN.....	124
3.1. UNIDADES DE ANÁLISIS.....	124
3.2. MUESTRA.....	124
3.2.1. Muestra cuantitativa.....	124
3.2.2. Muestra cualitativa.....	125
3.3. PERIODO DE LOS DATOS.....	125
3.4. TIPO Y NIVEL DE INVESTIGACIÓN.....	126
3.4.1. Tipo de investigación.....	126
3.4.2. Nivel de investigación.....	126
3.5. PERIODO DE ANÁLISIS.....	126
3.6. FUENTES DE INFORMACIÓN E INSTRUMENTOS UTILIZADOS	126
3.7. TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS	126
3.7.1. Técnicas.....	127
3.7.1.1. Minería de texto.....	127
3.7.1.2. Análisis de sentimiento.....	128
3.7.1.3. Encuestas.....	128
3.7.2. Instrumentos.....	128
3.7.2.1. API Graph.....	129
3.7.2.2. Facepager.....	129

3.7.2.3. Cuestionarios.....	129
CAPÍTULO IV.....	131
MARCO ORGANIZACIONAL.....	131
4.1. EMPRESA DE ESTUDIO .....	131
4.1.1. Visión .....	131
4.1.2. Misión.....	131
4.2. DESCRIPCIÓN DEL CONTEXTO .....	132
4.3. LÍNEAS DE NEGOCIO .....	133
4.3.1. Riesgos humanos.....	133
4.3.2. Riesgos corporativos.....	134
4.3.3. Grupos de afinidad .....	135
CAPÍTULO V.....	137
IMPLEMENTACIÓN DEL MODELO .....	137
5.1. PLANIFICACIÓN DEL PROYECTO .....	138
5.1.1. Objetivos del proyecto.....	138
5.1.2. Definición del proyecto .....	138
5.1.3. Alcance del proyecto.....	139
5.2. REQUERIMIENTOS DEL NEGOCIO .....	140
5.3. MODELAMIENTO DIMENSIONAL.....	140
5.3.1. Modelo dimensional .....	141

5.3.1.1.	Componentes del modelo dimensional.....	142
5.4.	MODELO DE INTELIGENCIA DE NEGOCIOS .....	146
5.5.	ARQUITECTURA DEL MODELO DE INTELIGENCIA DE NEGOCIOS.....	147
5.6.	HERRAMIENTAS DEL MODELO.....	148
5.6.1.	SQL Server Database Engine .....	148
5.6.1.1.	Descarga Microsoft SQL Server 2016 .....	149
5.6.1.2.	Instalar Microsoft SQL Server 2016.....	150
5.6.2.	SQL Server Data Tools (SSDT) .....	151
5.6.2.1.	Instalar SQL Server Data Tools (SSDT) .....	151
5.6.2.2.	SQL Server Analysis Services (SSAS), SQL Server Integration Services (SSIS) y SQL Server Reporting Services (SSRS)	152
5.6.3.	Facepager .....	155
5.6.3.1.	Descarga del Instalador Facepager.....	155
5.6.3.2.	Instalar Facepager en equipo cliente.....	155
5.7.	CONSTRUCCIÓN DEL MODELO .....	156
5.7.1.	Extracción, transformación y carga (ETL) .....	156
5.7.1.1.	Extracción de comentarios usando API Graph .....	157
5.7.1.1.1.	Registrar cuenta de usuario desarrollador.....	157
5.7.1.1.2.	Explorador de la API Graph.....	158

5.7.1.1.3.	Crear nueva aplicación (APP) .....	159
5.7.1.1.4.	Obtener token de acceso .....	161
5.7.1.1.5.	Extender caducidad token de acceso.....	162
5.7.1.1.6.	Acceder a datos publicos con Python.....	163
5.7.1.2.	Extracción de comentarios con Facepager.....	167
5.7.1.2.1.	Diseño de ventana principal de Facepager .....	167
5.7.1.2.2.	Definición de la base datos en Facepager .....	168
5.7.1.2.3.	Definición de nodos de recolección de datos .....	169
5.7.1.2.4.	Selección de módulos de extracción .....	170
5.7.1.2.5.	Exportar nodos a archivos .csv.....	174
5.7.2.	Conjunto de datos de entrenamiento .....	175
5.7.3.	Clasificación de texto .....	176
5.7.3.1.	Selección del algoritmo de clasificación .....	177
5.7.3.2.	Modelo de clasificación de texto .....	181
5.7.4.	Análisis de sentimiento.....	182
5.7.4.1.	Librerías del modelo .....	182
5.7.4.2.	Preparación de los datos .....	183
5.7.4.3.	Modelo de análisis de sentimiento.....	184
5.7.5.	Construcción del almacén de datos .....	185
5.7.5.1.	Data Staging Area (STG).....	186

5.7.5.2.	Operational Data Store (ODS) .....	186
5.7.5.3.	Datamart area (DWH).....	186
5.7.5.4.	Conexión al servidor de base datos.....	187
5.7.5.5.	Creación de carpetas del modelo .....	188
5.7.5.6.	Creación bases de datos del modelo.....	190
5.7.5.7.	Creación tablas del modelo .....	191
5.7.5.7.1.	Tablas data staging area .....	191
5.7.5.7.2.	Tablas Data Warehouse area.....	196
5.7.6.	Creación del proyecto Integration Services (SSIS) .....	201
5.7.6.1.	Agregar paquete de Integration Services al proyecto .....	202
5.7.7.	Creación del proyecto Analysis Services (SSAS) .....	204
5.7.7.1.	Acceso a la fuente de datos .....	205
5.7.7.2.	Vista de fuente de datos .....	208
5.7.7.3.	Creación de un nuevo cubo .....	211
5.7.8	Creación del tablero de control en Power BI .....	215
5.7.8.1.	Acceso a la fuente de datos .....	215
5.7.8.2.	Vista de la tabla de datos .....	217
5.7.8.3.	Vista del modelo de datos.....	218
5.7.8.4.	Tablero de control.....	219
CAPÍTULO VI.....		221

ANÁLISIS ECONOMICO FINANCIERO .....	221
6.1. INVERSIÓN DEL PROYECTO .....	221
6.2. RESUMEN DE LOS COSTOS DEL PROYECTO .....	223
6.3. RENTABILIDAD DEL PROYECTO.....	224
CAPÍTULO VII.....	226
ANÁLISIS Y DISCUSIÓN DE RESULTADOS .....	226
7.1. RESULTADOS DE LA INVESTIGACIÓN .....	226
7.1.1. Extracción de comentarios .....	226
7.1.2. Clasificación de texto .....	228
7.1.3. Análisis de sentimiento.....	229
7.1.4. Tablero de control .....	233
7.1.5. Indicadores del tablero de control .....	237
7.1.5.1. Cantidad de comentarios por publicación .....	238
7.1.5.2. Cantidad de comentarios por mes .....	239
7.1.5.3. Cantidad de comentarios por tema.....	240
7.1.5.4. Cantidad de comentarios por sentimiento .....	241
7.1.6. Ventajas .....	242
7.2. HERRAMIENTAS DE INVESTIGACIÓN .....	244
7.3. PRUEBA DE HIPÓTESIS.....	245
7.3.1. Prueba de hipótesis general.....	245

7.3.1.1.	Hipótesis nula ( $H_0$ ) .....	245
7.3.1.2.	Hipótesis alternativa ( $H_\alpha$ ) .....	245
7.3.1.3.	Encuesta realizada .....	245
7.3.1.4.	Matriz de frecuencia observada (FO) .....	246
7.3.1.5.	Matriz de frecuencia esperada (FE).....	247
7.3.1.6.	Cálculo del chi-cuadrado .....	247
7.3.1.7.	Prueba de bondad de ajuste.....	249
7.3.2.	Prueba de la primera hipótesis.....	249
7.3.3.	Prueba de la segunda hipótesis .....	250
7.3.4.	Prueba de la tercera hipótesis.....	251
7.3.5.	Prueba de la cuarta hipótesis.....	252
7.3.5.1.	Hipótesis nula ( $H_0$ ) .....	252
7.3.5.2.	Hipótesis alternativa ( $H_a$ ) .....	252
7.3.5.3.	Encuesta realizada .....	252
7.3.5.4.	Matriz de frecuencia observada ( $F_o$ ) .....	253
7.3.5.5.	Matriz de frecuencia esperada ( $F_E$ ) .....	253
7.3.5.6.	Cálculo del estadístico chi-cuadrado .....	254
7.3.5.7.	Prueba de bondad de ajuste.....	255
7.4.	DISCUSIÓN DE RESULTADOS.....	256
	CONCLUSIONES .....	260

RECOMENDACIONES.....	262
GLOSARIO.....	264
REFERENCIAS BIBLIOGRÁFICAS.....	268
ANEXOS.....	283

## ÍNDICE DE TABLAS

Tabla N° 1: Matriz de operacionalización de variables .....	41
Tabla N° 2: Matriz de consistencia.....	42
Tabla N° 3: Cuadro de antecedentes de la investigación .....	51
Tabla N° 4: Tabla de hechos: FACT_FACEBOOK .....	143
Tabla N° 5: Dimensión DIM_TEMA.....	143
Tabla N° 6: Dimensión DIM_SENTIMIENTO .....	143
Tabla N° 7: Dimensión DIM_FECHA .....	144
Tabla N° 8: Dimensión DIM_PAGINA.....	144
Tabla N° 9: Dimensión DIM_PUBLICACION .....	144
Tabla N° 10: Dimensión DIM_COMENTARIO .....	145
Tabla N° 11: Resultados test algoritmo Naive Bayes.....	178
Tabla N° 12: Resultados test algoritmo Decision Tree.....	178
Tabla N° 13: Resultados test algoritmo SVM.....	179
Tabla N° 14: Bases de datos de la solución .....	185
Tabla N° 15: Características data staging area .....	186
Tabla N° 16: Base datos operational data store (ODS) .....	186
Tabla N° 17: Características base datos.....	187
Tabla N° 18: Carpetas del modelo .....	188
Tabla N° 19: Costos del proyecto - recursos humanos.....	222
Tabla N° 20: Costos del proyecto - recurso de hardware .....	223
Tabla N° 21: Costos del proyecto - recurso de software.....	223
Tabla N° 22: Resumen de costos del proyecto .....	223

Tabla N° 23: Cálculo del valor presente de los flujos (en miles de nuevos soles) .....	225
Tabla N° 24: Escala likert de tipo frecuencia .....	244
Tabla N° 25: Encuesta 01 – hipótesis general .....	246
Tabla N° 26: Matriz de frecuencia observada (Fo).....	246
Tabla N° 27: Matriz de frecuencia esperada (Fe) .....	247
Tabla N° 28: Matriz de cálculo del chi cuadrado.....	248
Tabla N° 29: Matriz de cálculo de P-value .....	248
Tabla N° 30: Encuesta 02 – cuarta hipótesis .....	253
Tabla N° 31: Matriz de frecuencia observada (Fo).....	253
Tabla N° 32: Matriz de frecuencia esperada (Fe) .....	254
Tabla N° 33: Matriz de cálculo del chi cuadrado.....	255
Tabla N° 34: Matriz de cálculo de P-value .....	255

## ÍNDICE DE FIGURAS

Figura N° 1: Evolución de la Inversión Anual, 2014 - 2020 (En Millones de dólares).....	22
Figura N° 2: Comparativo de Formatos, 2018 - 2020 (Porcentaje).....	23
Figura N° 3: Población Censada Según Condición de Tenencia de Seguros de Salud, 2007 y 2017 (Absoluto y Porcentaje).....	23
Figura N° 4: Población Censada sin Seguro de Salud, Según Grupos de Edad, 2017 (Absoluto y porcentaje).....	24
Figura N° 5: Población Censada sin Seguro de Salud, Según Ciclo de Vida, 2017 (Porcentaje respecto al total de población censada por grupo de edad).....	25
Figura N° 6: Población Censada de 15 y más años, sin seguro de salud, según nivel educativo, 2017 (Absoluto y porcentaje).....	26
Figura N° 7: Población ocupada con seguro de salud, según área de residencia, 2007, 2016, 2017 (Miles de personas y porcentajes).....	27
Figura N° 8: Población ocupada con seguro de salud, 2007-2017 (porcentaje del total de ocupados).....	28
Figura N° 9: Acceso a Seguro de Salud .....	29
Figura N° 10: Población ocupada por empleo formal e informal, según ámbito geográfico 2008, 2016 y 2017 (miles de personas).....	30
Figura N° 11: Población ocupada en condición de trabajador independiente por sexo, según acceso a otras formas de servicio financiero no formales, 2017 (Porcentaje) .....	31

Figura N° 12: Diagrama de Ishikawa .....	36
Figura N° 13: Empresas de Seguros Autorizadas (Al 31 de diciembre del 2020).....	53
Figura N° 14: Ranking de Primas de Seguros Netas por Ramos Al 31 de diciembre del 2020 (En Miles de Soles).....	54
Figura N° 15: Estructuras de las Primas de Seguros Netas, por Ramos Al 31 de diciembre del 2020 (En porcentajes) .....	55
Figura N° 16: Evolución de las Primas de Seguros Netas por Ramos Al 31 de diciembre del 2020.....	55
Figura N° 17: Primas Según Ramos Al 31 de diciembre del 2020 (En Miles de Soles).....	56
Figura N° 18: Variación anual de Primas de Seguros Netas, por Ramos (En Porcentaje).....	57
Figura N° 19: Primas de Seguros Netas de Seguros del SPP (En millones de Soles).....	59
Figura N° 20: Información Financiera de Corredores de Seguros Al 31 de diciembre de 2019 (En Soles).....	60
Figura N° 21: Funciones de un Sistema de Información.....	63
Figura N° 22: Dimensiones de un Sistema de Información.....	64
Figura N° 23: Cómo los Sistemas de Información Obtienen datos del TPS	67
Figura N° 24: Arquitectura Inteligencia de Negocios (BI).....	70
Figura N° 25: Reemplazando claves naturales o de negocio por claves sustitutas.....	77
Figura N° 26: Esquema Estrella vs cubos OLAP .....	78

Figura N° 27: Tabla de Hechos y Dimensiones relacionadas en un Esquema Estrella.....	78
Figura N° 28: Arquitectura Inteligencia de Negocios Social (SBI).....	83
Figura N° 29: Machine Learning Overview .....	84
Figura N° 30: Aprendizaje Supervisado .....	86
Figura N° 31: Aprendizaje del Support Vector Classifiers.....	89
Figura N° 32: Aprendizaje No Supervisado .....	90
Figura N° 33: Los 3 pasos básicos en la Indexación de Texto .....	96
Figura N° 34: Proceso de Tokenización.....	97
Figura N° 35: Proceso de Tokenización de texto.....	98
Figura N° 36: Proceso de Stemming.....	99
Figura N° 37: Clasificador Basado en Reglas para la categoría de Trigo ..	103
Figura N° 38: Representación de Clasificadores de Texto .....	105
Figura N° 39: Proceso de Minería de Texto .....	113
Figura N° 40: Índice de Promotores Netos de una empresa.....	116
Figura N° 41: Estructura de una API.....	117
Figura N° 42: Ciclo de Vida del Proyecto.....	138
Figura N° 43: Modelo Dimensional .....	142
Figura N° 44: Modelo de Inteligencia de Negocios .....	146
Figura N° 45: Arquitectura del Modelo de Inteligencia de Negocios .....	147
Figura N° 46: Descargar Microsoft SQL Server 2016 .....	150
Figura N° 47: Instalar Microsoft SQL Server 2016.....	151
Figura N° 48: Instalar SSDT con Visual Studio.....	152
Figura N° 49: Instalar SSDT con soporte para SSAS, SSIS, SSRS .....	153

Figura N° 50: Instalación de SSDT .....	154
Figura N° 51: Descarga del Instalador Facepager .....	155
Figura N° 52: Registrar Cuenta de Usuario Desarrollador .....	158
Figura N° 53: Facebook para Desarrolladores.....	158
Figura N° 54: Explorador de la API Graph .....	159
Figura N° 55: Crear Nueva Aplicación (APP).....	160
Figura N° 56: Crear Nuevo Identificador de Aplicación (APP) .....	160
Figura N° 57: Obtener Token de Acceso .....	162
Figura N° 58: Extender Caducidad del Token de Acceso .....	163
Figura N° 59: Instalar Paquete Facebook-sdk .....	164
Figura N° 60: Acceder a datos públicos de Facebook con Python .....	165
Figura N° 61: Respuesta datos básicos en notación Json.....	165
Figura N° 62: Respuesta Likes en notación Json .....	166
Figura N° 63: Respuesta Posts en notación Json.....	166
Figura N° 64: Diseño de la Ventana Principal en Facepager.....	167
Figura N° 65: Base Datos del Proyecto en Facepager .....	168
Figura N° 66: Ventana de Creación de Base Datos en Facepager.....	169
Figura N° 67: Ventana de Definición de Nodos de Recolección de Datos. 169	
Figura N° 68: Ventana de Definición de Nodos de Recolección de Datos. 170	
Figura N° 69: Módulos de Recolección de Datos.....	171
Figura N° 70: Presets de Extracción de Datos.....	172
Figura N° 71: Configuración Consulta Extracción de Comentarios.....	173
Figura N° 72: Comentarios Retornados por a API Graph .....	173
Figura N° 73: Exportar Nodos a Archivo CSV.....	174

Figura N° 74: Conjunto de Datos de Entrenamiento .....	176
Figura N° 75: Comparación de Precisión algoritmos: Naive Bayes, Decision Tree y SVM.....	179
Figura N° 76: Modelo de Clasificación de Texto .....	181
Figura N° 77: Librerías del Modelo Análisis de Sentimiento .....	182
Figura N° 78: Carga del Conjunto de Datos de Entrenamiento .....	183
Figura N° 79: Procesamiento del Conjunto de Datos de Entrenamiento ...	183
Figura N° 80: Modelo de Análisis de Sentimiento .....	184
Figura N° 81: Conexión al Servidor de Base Datos .....	188
Figura N° 82: Estructura de carpetas base datos: MBIS-STG-01 .....	189
Figura N° 83: Estructura de carpetas base datos: MBIS-ODS-01.....	189
Figura N° 84: Estructura de carpetas base datos: MBIS-DWH-01 .....	189
Figura N° 85: Script creación de base datos [MBIS-STG-01] .....	190
Figura N° 86: Script creación de base datos [MBIS-ODS-01].....	190
Figura N° 87: Script creación de base datos [MBIS-DWH-01].....	190
Figura N° 88: Script tabla FACEBOOK_000 .....	191
Figura N° 89: Script tabla FACEBOOK_001 .....	192
Figura N° 90: Script tabla FACEBOOK_002 .....	192
Figura N° 91: Script tabla FACEBOOK_003 .....	193
Figura N° 92: Script tabla FACEBOOK_004 .....	194
Figura N° 93: Script tabla FACEBOOK_001_ERROR .....	195
Figura N° 94: Diagrama de tablas Data Staging Area.....	196
Figura N° 95: Script tabla de hechos .....	197
Figura N° 96: Script tabla de dimensión: DIM_FECHA .....	197

Figura N° 97: Script tabla de dimensión: DIM_TEMA .....	198
Figura N° 98: Script tabla de dimensión: DIM_SENTIMIENTO.....	198
Figura N° 99: Script tabla de dimensión: DIM_PAGINA.....	198
Figura N° 100: Script tabla de dimensión: DIM_PUBLICACION.....	199
Figura N° 101: Script tabla de dimensión: DIM_COMENTARIO.....	199
Figura N° 102: Script restricciones base de datos .....	200
Figura N° 103: Selección de plantilla de proyecto Integration Services .....	201
Figura N° 104: Carpetas plantilla de proyecto Integration Services .....	202
Figura N° 105: Agregar nuevo paquete Integration Services .....	203
Figura N° 106: Flujo de control paquete Integration Services (SSIS) .....	203
Figura N° 107: Creación de un proyecto Analysis Services.....	204
Figura N° 108: Selección de plantilla de proyecto Analysis Services.....	205
Figura N° 109: Carpetas de plantilla de proyecto Analysis Services.....	205
Figura N° 110: Creación de una nueva fuente de datos .....	206
Figura N° 111: Ventana asistente de fuente de datos.....	206
Figura N° 112: Administrador de conexiones.....	207
Figura N° 113: Ventana explorador de soluciones.....	208
Figura N° 114: Crear nueva vista de fuente de datos .....	208
Figura N° 115: Ventana vista fuente datos, seleccionar fuente datos.....	209
Figura N° 116: Ventana vista fuente datos, seleccionar tablas y vistas .....	210
Figura N° 117: Diseñador de vista fuente de datos .....	211
Figura N° 118: Creación de un nuevo Cubo .....	211
Figura N° 119: Ventana asistente de cubo, Seleccionar tablas de medidas de grupo.....	212

Figura N° 120: Ventana asistente de cubo, Seleccionar dimensiones.....	213
Figura N° 121: Diseñador del cubo .....	214
Figura N° 122: Editor de reportes en Power BI .....	215
Figura N° 123: Obtener acceso a la fuente de datos .....	216
Figura N° 124: Ventana SQL Server Analysis Services Database .....	216
Figura N° 125: Ventana de navegación del cubo.....	217
Figura N° 126: Vista tabular de la tabla de datos.....	218
Figura N° 127: Vista del modelo de datos.....	219
Figura N° 128: Tablero de control en Power BI .....	220
Figura N° 129: Tuit originó la debacle.....	231
Figura N° 130: Tuit aclaración BMW USA .....	232
Figura N° 131: Cuadrante Mágico Gartner para plataformas de Inteligencia de Negocios y Analítica .....	234
Figura N° 132: Arquitectura de flujo de datos de Bayer .....	236
Figura N° 133: Script MDX - Cantidad de comentarios por publicación.....	238
Figura N° 134: Script MDX - Cantidad de comentarios por mes.....	239
Figura N° 135: Script MDX - Cantidad de comentarios por tema.....	240
Figura N° 136: Script MDX - Cantidad de comentarios por sentimiento ....	241
Figura N° 137: Cálculo porcentaje integración extracción de comentarios	249
Figura N° 138: Comparativo precisión algoritmos de clasificación .....	250
Figura N° 139: Porcentaje precisión modelo Análisis de Sentimiento.....	251

## **RESUMEN**

La presente investigación se basa principalmente en la implementación de un modelo de Inteligencia de Negocios basado en minería de opinión sobre comentarios a publicaciones expresados por usuarios en redes sociales. El modelo será implementado en una empresa corredora de seguros en el Perú.

Uno de los grandes desafíos que enfrentan las empresas en la nueva era de la información es tener la capacidad de transformar los datos en información y convertir la información en conocimiento para la toma de decisiones.

Asimismo, En los últimos años las empresas enfrentan nuevos desafíos como son: la innovación, el incremento de uso de internet, las redes sociales, los teléfonos inteligentes, la conectividad y la ubicuidad de las personas. Para lograr el éxito en la innovación es fundamental una profunda comprensión de los clientes.

Existen tres motivaciones para que las personas estén dispuestos a recibir y transmitir contenido de una empresa o marca en redes sociales: social, emocional y funcional. Estas motivaciones se basan principalmente

en, como enviar señales a otros sobre la experiencia, la singularidad o el estatus social de uno, compartir sentimientos positivos o negativos sobre las marcas y la tendencia de obtener y proporcionar información a otros.

En consecuencia, las empresas tienden a invertir más recursos en publicidad en redes sociales lo cual genera que las publicaciones de las empresas lleguen a más usuarios en un mercado mucho más segmentado, lo cual permite a los usuarios reaccionar a través de sus opiniones y emociones expresadas en comentarios en redes sociales.

Según el informe de inversión publicitaria del IAB Perú, a pesar del contexto adverso que enfrentamos como es la pandemia del coronavirus, la inversión publicitaria digital en el Perú creció un 6% en el 2020. En este informe se destaca el crecimiento de un +6% el formato de Video y un +7% en el formato Social Ads (publicidad en redes sociales).

Los datos no estructurados están dominando a los datos estructurados con un volumen de hasta un 80% en comparación con su contraparte, los datos estructurados con solo un 20%.

Según reporte del portal Statista.com en el 2021, en la plataforma de red social Facebook existen 2,79 billones de usuarios que se conectan mensualmente, también es una de las plataformas de red social que tiene mayor interacción de sus usuarios, según cifras del portal Zephoria Digital Marketing en el 2019 cada 60 segundos se publican 510,000 comentarios, se actualizan 293,000 estados y se suben 136,000 fotos.

Asimismo, según cifras del portal wearesocial.com & hootsuite.com, en el Perú existen 24 millones de usuarios activos en redes sociales lo cual representa un 73% de la población. 23 millones de estos usuarios activos utilizan un teléfono inteligente para acceder a internet, lo cual representa un 70% de la población.

Del mismo modo, en el Perú existen 39.08 millones de dispositivos móviles, lo cual demuestra que existen más celulares que habitantes, esto representa un 119% como porcentaje de la población.

Respecto al mercado asegurador, de acuerdo con las cifras estadísticas de la Superintendencia Nacional de Banca Seguros y AFP al 31 de diciembre del 2020, existen 18 empresas de seguros, las cuales se agrupan en 3 categorías, de acuerdo con los ramos en los cuales están enfocadas. Siete empresas están dedicadas a Ramos Generales y de Vida, siete dedicadas exclusivamente a Ramos Generales, y cuatro se dedican exclusivamente a Ramos de Vida.

La participación de mercado del sector asegurador está compuesta de las siguientes empresas: Rímac con un 30.64%, Pacífico Seguros con un 26.06%, La Positiva con 8.15% y Mapfre Perú con un 8.14%; las cuales son las empresas de seguros más reconocidas en el sector asegurador en el Perú.

La producción de Primas de Seguros Netas (PSN) anualizadas ascendieron a S/ 14,021,4 millones, -0.7% puntos porcentuales con respecto

a diciembre de 2019. La composición de la producción estuvo liderada por los Ramos Generales con S/ 5,340,241 y una participación de 38.1%, +1.3 puntos porcentuales a la registrada en diciembre de 2019. A su vez, los Seguros de Vida con S/ 4,722,209 y una participación de 33,7%, -0,4 puntos porcentuales con respecto s diciembre 2019.

Las empresas de seguros se enfrentan al desafío de diseñar productos y servicios más competitivos que sean apoyados por la tecnología para poder ofrecer a sus clientes la reducción en el pago de su prima de seguro. Según el informe de Forbes, manejar de forma segura le puede ahorrar a los usuarios de seguros hasta un 30% en el pago de sus primas de seguros.

La presente investigación de Inteligencia de Negocios permitirá a los ejecutivos de operaciones, ejecutivos de siniestros, gerencia general, gerencia comercial y gerencia de operaciones a poder analizar las opiniones y emociones expresadas a través de comentarios en redes sociales, para tomar decisiones objetivas basada en datos, respecto al nivel de satisfacción que tienen usuarios acerca de los servicios entregados por la empresa lo cual contribuirá a la mejora del servicio al cliente.

La fuente de información de la presente investigación son los comentarios a las publicaciones de empresas del sector asegurador que se encuentran disponibles en las plataforma de red social: Facebook, los cuales serán extraídos a través de programas rastreadores y servicios API, que aplicando algoritmos de clasificación de texto serán clasificados a una

categoría y a través de técnicas de Análisis de Sentimiento se identificarán patrones de comportamiento y formas de expresión que tienen los usuarios al reaccionar de manera positivas, negativa o neutra acerca de un producto o servicio.

En la presente investigación se encontrará conceptos técnicos relacionados a: Inteligencia de Negocios, Aprendizaje de Máquina, Minería de Texto, Clasificación de Texto, Análisis de Sentimiento, Servicios RESTful API.

Adicionalmente, se podrá encontrar los procedimientos para poder integrar estos conceptos a través de componentes de software y hardware logrando la sincronización en su funcionamiento para luego implementar proyectos en forma concreta que se integren a los procesos de negocio de una empresa del sector asegurador y que permita a los empleados a tomar decisiones en forma acertada acerca del comportamiento que tienen los usuarios en redes sociales, permitiendo desarrollar estrategias para mejorar el nivel de servicio al cliente.

**Palabras clave:** Inteligencia de Negocios, Aprendizaje de Máquina, Minería de Texto, Minería de Opinión, Clasificación de Texto, Análisis de Sentimiento, Support Vector Machine (SVM), RESTful API, Corredores de Seguros, Facebook.

## **ABSTRACT**

This research is mainly based on the implementation of a Business Intelligence model based on opinion mining on comments to publications expressed by users on social networks. The model will be implemented in an insurance brokerage company in Peru.

One of the great challenges that companies face in the new information age is having the ability to transform data into information and convert information into knowledge for decision making.

Likewise, in recent years companies face new challenges such as: innovation, increased use of the Internet, social networks, smartphones, connectivity, and the ubiquity of people. To achieve success in innovation, a deep understanding of customers is essential.

There are three motivations for people to be willing to receive and transmit content from a company or brand on social networks: social, emotional, and functional. These motivations are primarily based on, such as sending signals to others about one's expertise, uniqueness, or social status, sharing positive or negative feelings about brands, and the tendency to obtain and provide information to others.

Consequently, companies tend to invest more resources in advertising on social networks, which means that company publications reach more users in a much more segmented market, which allows users to react through their opinions and emotions expressed. in comments on social networks.

According to the IAB Peru advertising investment report, despite the adverse context we are facing such as the coronavirus pandemic, digital advertising investment in Peru grew by 6% in 2020. This report highlights the growth of +6 % the Video format and +7% in the Social Ads format (advertising on social networks).

Unstructured data is dominating structured data with up to 80% volume compared to its counterpart, structured data at just 20%.

According to a report from the Statista.com portal in 2021, on the Facebook social network platform there are 2.79 billion users who connect monthly, it is also one of the social network platforms that has the greatest interaction from its users, according to figures from the Zephoria Digital Marketing portal in 2019 every 60 seconds 510,000 comments are posted, 293,000 statuses are updated, and 136,000 photos are uploaded.

Likewise, according to figures from the portal wearesocial.com & hootsuite.com, in Peru there are 24 million active users on social networks, which represents 73% of the population. 23 million of these active users use a smartphone to access the internet, which represents 70% of the population.

According to the statistical figures of the Superintendencia Nacional de Banca Seguros y AFP as of June 2017, there are 22 insurance companies in Peru. Insurance companies can be grouped into 3 categories, according to the branches in which they are focused.

Regarding the insurance market, according to the statistical figures of the Superintendencia Nacional de Banca Seguros y AFP as of December 31, 2020, there are 18 insurance companies, which are grouped into 3 categories, according to the branches in which they are focused. Seven companies are dedicated to General and Life risks, seven exclusively dedicated to General risks, and four are exclusively dedicated to Life risks.

The market share of the insurance sector is made up of the following companies: Rímac with 30.64%, Pacífico Seguros with 26.06%, La Positiva with 8.15% and Mapfre Perú with 8.14%, which are the most recognized insurance companies in the insurance sector in Peru.

The annualized production of Net Insurance Premiums (PSN) amounted to S/ 14,021.4 million, -0.7% percentage points compared to December 2019. The composition of the production was led by the General risks with S/ 5,340,241 and a participation of 38.1%, +1.3 percentage points to that registered in December 2019. In turn, Life Insurance with S/ 4,722,209 and a participation of 33.7%, -0.4 percentage points compared to December 2019.

Insurance companies face the challenge of designing more competitive products and services that are supported by technology in order to offer their clients reduced insurance premium payments. According to the Forbes report, driving safely can save insurance users up to 30% on the payment of their insurance premiums.

This Business Intelligence research will allow operations executives, claims executives, general management, commercial management and operations management to be able to analyze the opinions and emotions expressed through comments on social networks, to make objective decisions based on data, regarding the level of satisfaction that users have about the services delivered by the company, which will contribute to the improvement of customer service.

The source of information for this research are the comments on the publications of companies in the insurance sector that are available on the social network platform: Facebook, which will be extracted through crawler programs and API services, which apply classification algorithms. of text will be classified into a category and through Sentiment Analysis techniques, behavior patterns and forms of expression that users have when reacting positively, negatively or neutrally about a product or service will be identified.

In the present investigation, technical concepts related to: Business Intelligence, Machine Learning, Text Mining, Text Classification, Sentiment Analysis, RESTful API Services will be found.

Additionally, it will be possible to find the procedures to be able to integrate these concepts through software and hardware components, achieving synchronization in their operation to later implement projects in a concrete way that are integrated into the business processes of a company in the insurance sector and that allows employees to make correct decisions about the behavior of users on social networks, allowing the development of strategies to improve the level of customer service.

**Keywords:** Business Intelligence, Machine Learning, Text Mining, Opinion Mining, Text Classification, Sentiment Analysis, Support Vector Machine (SVM), Insurance Broker, Facebook.

## **INTRODUCCIÓN**

En los últimos años las empresas han tenido que enfrentar grandes desafíos debido a los cambios tecnológicos, como son la conectividad y el fácil acceso a Internet, lo cual está originando que la cantidad de usuarios que hacen uso de las redes sociales para comunicarse y compartir información siga incrementándose mes a mes. Según cifras estadísticas de Facebook al 31 de diciembre del 2019, existen más de 2,79 billones de usuarios que se conectan mensualmente (MAU).

La plataforma de red social Facebook es una de las que tienen la mayor cantidad de interacción de sus usuarios. Cada 60 segundos se publican 510,000 comentarios, se actualizan 293,000 estados y se suben 136,000 fotos. (Zephoria.com, Zephoria Digital Marketing, 2019)

Las redes sociales brindan a las empresas una forma de comunicación directa con sus consumidores para realizar una campaña publicitaria o comunicarse directamente con ellos y conocer la perspectiva que tienen acerca de sus productos o servicios.

En redes sociales y en la comunicación tradicional el mensaje sigue siendo el factor clave para lograr una comunicación exitosa. Para comunicar

a un mercado objetivo en redes sociales, se debe diseñar un mensaje claro y directo que haga que los usuarios reaccionen y los motive a comentar y compartir el mensaje a sus contactos.

De acuerdo con las investigaciones de (Lovett, Peres, & Shachar, 2013) existen tres motivaciones: social, emocional y funcional, para que los consumidores estén dispuestos a recibir y transmitir contenido de una marca. Estas motivaciones o impulsores se deben al deseo de enviar señales a otros sobre su experiencia, estatus social, compartir sentimientos positivos o negativos y la disposición que tienen algunos de compartir información con otros para beneficio de los demás.

Los comentarios de los consumidores en redes sociales son de naturaleza subjetiva y están cargados de emociones y sentimientos. Dichos comentarios se encuentran disponibles en las plataformas de redes sociales. Para poder transformarlos en puntos de vista coherente y significativo acerca de productos y servicios, se necesitan herramientas de extracción, clasificación de texto y de análisis de sentimiento.

## **CAPÍTULO I**

### **PROBLEMA DE INVESTIGACIÓN**

#### **1.1. ANTECEDENTES DE LA INVESTIGACIÓN**

##### **1.1.1. Toma de decisiones**

Toda organización necesita de información para la toma de decisiones, el conocimiento es poder, es una frase que se le atribuye a Francis Bacon en su obra “scientia potentia est”, (Sir Francis Bacon, 1668)

El conocimiento es adquirido en función de definir una verdad. En una sociedad, la función de los que definen la verdad es la transmisión de este conocimiento que se hace mediante normas y conductas. Por lo tanto, en una sociedad ejercer conocimiento es sinónimo del ejercicio del poder. (Foucault, 1977)

Lo que significa que mientras más conocimiento se tenga sobre algo o alguien, más poder tendrá, el conocimiento sobre alguna materia nos da más opciones y formas de enfrentar las oportunidades y amenazas del entorno.

Las empresas que sean capaces de transformar datos en información y conocimiento pueden utilizarlos para tomar decisiones más rápidas y efectivas y así lograr una ventaja competitiva. (Vercellis, 2009)

Las buenas decisiones llevan a la selección de metas y líneas de acción apropiadas que mejoran el rendimiento organizacional, mientras las malas elecciones afectan negativamente el desempeño de una organización. (Jones & George , 2018)

En proceso de toma de decisiones grupal de atributos múltiples, los tomadores de decisiones son personas que trabajan en diferentes áreas y tienen diferentes especialidades. La forma en que expresan su juicio depende tanto de la naturaleza de las características que describen las alternativas como de su propia cognición. Además, se debe asumir que, con cognición limitada, falta de información y tiempo adecuados. Es muy difícil que un tomador de decisiones esté familiarizado con todos los aspectos del problema. (Chen & Zou, 2019)

Existen dos estilos para la toma de decisiones: basadas en la intuición las cuales son tomadas en la sociedad, la política y la economía, ahora son reemplazadas por decisiones basadas en datos, que son más razonables. (Jeong, Ko, Son, & Yoon, 2020)

### **1.1.2. Inteligencia de Negocios**

Los sistemas de inteligencia de negocios tienden a promover un enfoque científico y racional para la gestión de empresas y organizaciones complejas.

Incluso el uso de una hoja de cálculo electrónica requiere por parte de los tomadores de decisiones una representación mental de los flujos financieros. (Vercellis, 2009)

La inteligencia de negocios social combina datos de las organizaciones con contenido generado por el usuario (CGU) para que esta manera los tomadores de decisiones estén al tanto de las tendencias provenientes del entorno. Un papel clave en el análisis de CGU textual lo juegan los temas, entendidos como conceptos específicos de interés dentro de un área temática. (Gallinucci, Golfarelli, & Rizzi, 2015)

Para brindar información a las diferentes unidades de negocio y a los niveles: estratégicos, táctico y operativo de la organización, en la administración de negocios se hace necesario contar con herramientas de visualización en forma gráfica. Según los autores (Killen, Gerald, & Kock, 2020), debido a la importancia que tiene la información gráfica en la comunicación y en la toma de decisiones, existe un interés creciente de los directivos en mostrar información a través de visualizaciones en forma gráfica como son los: gráficos, dibujos, diagramas de flujo lo cual permitirá comunicar, persuadir e informar de manera más efectiva.

### **1.1.3. Redes sociales**

En los últimos años, las organizaciones enfrentan nuevos desafíos que antes no existían, como son: la innovación, el incremento de uso de internet, las

redes sociales, los teléfonos inteligentes, la conectividad y la ubicuidad de las personas.

En los últimos años, la mayoría de nosotros podemos conectarnos a Internet desde cualquier lugar, en cualquier momento y enviar mensajes continuamente a una comunidad virtual centrada en blogs, foros, redes sociales y similares. Esto ha resultado en la acumulación de enormes cantidades de contenido generado por usuarios (CGU), que incluyen geolocalización, preferencias, opiniones, noticias, etc. (Gallinucci, Golfarelli, & Rizzi, 2015)

En el escenario actual, las redes sociales juegan un papel relevante al brindar información sobre cualquier producto a partir de diferentes reseñas, blogs y comentarios. Para obtener información significativa de los sentimientos de las personas, académicos y profesionales aplican diferentes técnicas de aprendizaje automático. (Tripathy, Agrawal, & Rath, 2016)

El auge de la tecnología y los medios digitales brinda a los especialistas en marketing muchas oportunidades nuevas para llegar e influir en sus audiencias para alentar la compra. Sin embargo, esto significa que la planificación y la gestión de las comunicaciones de marketing son ahora más complejas que nunca. (Smartinsights, 2020)

El advenimiento de las tecnologías de almacenamiento de datos de bajo costo y la amplia disponibilidad de conexiones a internet han facilitado que

las personas y las organizaciones accedan a grandes cantidades de datos. (Vercellis, 2009)

La tecnología ofrece la plataforma de redes sociales que permite la comunicación directa con sus consumidores, de acuerdo con (SmartInsights, 2020) las tecnologías emergentes facilitan la satisfacción de las necesidades de los clientes y la creación de oportunidades en entornos B2B y B2C.

Según (Sarlan, Nadam, & Basri, 2014) las redes sociales brindan a las empresas una oportunidad de comunicarse directamente con sus clientes para realizar una campaña publicitaria o hablar directamente con ellos, así como conectarse con la perspectiva del cliente acerca sus productos y servicios.

Los datos de las redes sociales son claramente la base de evidencia más grande, rica y dinámica del comportamiento humano, lo que brinda nuevas oportunidades para comprender a las personas, los grupos y la sociedad. Los científicos innovadores y los profesionales de la industria encuentran cada vez más formas novedosas de recopilar, combinar y analizar automáticamente esta gran cantidad de datos. (Batinca & Treleaven, 2015)

La minería de las redes sociales es verdaderamente potencial y útil para extraer más información y obtener un conocimiento más profundo sobre los clientes. Se puede utilizar para identificar al cliente influyente, detectar grupos implícitos u ocultos, percibir la opinión de los clientes relacionada con la satisfacción de su producto o servicio, desarrollar sistemas de

recomendación para mantener a los clientes existentes y ganar nuevos, o construir y fortalecer la confianza entre los clientes o entre los clientes y otras partes interesadas. (Tundjungsari, 2013)

Las redes sociales se diferencian de la comunicación tradicional debido a sus características únicas y dinámicas, como la facilidad de participación de los usuarios. Sin embargo, el mensaje en sí sigue siendo un factor clave para una comunicación exitosa. (Sasser, Kilgour, & Hollebeek, 2014, págs. 29-52)

La capacidad de adaptarse a los cambios en las preferencias de los consumidores y de los canales sin dejar de ser relevante es fundamental, se necesita una investigación del público objetivo para revelar quiénes son sus clientes, cómo se comportan y qué plataformas prefieren usar. (SmartInsights, 2020)

La actividad de los usuarios proporciona una comprensión sustancial de la conducta, las experiencias, los puntos de vista y los intereses individuales. Existe una gran opción para agregar nuevas cualidades basadas en la personalidad a las interfaces de usuario. (Salloum, Al-Emran, & Shaalan, 2017)

Así mismo, existen tres motivaciones para que los consumidores estén dispuestos a recibir y transmitir contenido de una empresa o marca: social, emocional y funcional. Estas motivaciones se basan principalmente en necesidades, como enviar señales a otros sobre la experiencia, la

singularidad o el estatus social de uno, compartir sentimientos positivos o negativos sobre las marcas y la tendencia de obtener y proporcionar información a otros. (Lovett, Peres, & Shachar, 2013)

Esta enorme cantidad de datos sobre los gustos, pensamientos y acciones de las personas está suscitando un interés creciente por parte de los tomadores de decisiones porque les puede dar una percepción fresca y oportuna del estado de ánimo del mercado; además, muchas veces la difusión de los CGU (Contenido Generado por el Usuarios) está tan extendida que influye directamente de manera decisiva en los fenómenos de la empresa y la sociedad. (Gallinucci, Golfarelli, & Rizzi, 2015)

Dichos datos suelen ser heterogéneos en origen, contenido y representación, ya que incluyen transacciones comerciales, financieras y administrativas, rutas de navegación web, correos electrónicos, textos e hipertextos, y los resultados de pruebas clínicas, por nombrar solo algunos ejemplos. (Vercellis, 2009)

#### **1.1.4. Innovación de productos**

El éxito de la innovación se basa en una profunda comprensión de los clientes, conocer su entorno, sus rutinas diarias, sus preocupaciones y aspiraciones. (Osterwalder & Pigneur, 2011). El desafío que enfrentan las empresas de hoy consiste en conocer y segmentar perfectamente a su mercado objetivo para luego diseñar y dirigir una propuesta de valor del

modelo de negocio que cubra y supere las expectativas de los consumidores.

Las expectativas de valor de los clientes son más complejas que nunca. Además de estar mejor informados en general, quieren una conexión con las marcas que vaya más allá de los productos de calidad y los precios competitivos. (SmartInsights, 2020)

En el ámbito del diseño de productos y servicios, son varias empresas que recurren a sociólogos expertos para comprender mejor el panorama social. Este enfoque puede propiciar la aparición de modelos de negocios nuevos o mejores. (Osterwalder & Pigneur, 2011)

Debido a que cada vez más en la mayor parte de las industrias la oferta supera a la demanda, competir en esos mercados no es suficiente para lograr un alto desempeño de las organizaciones, las compañías deben ir más allá de la competencia con el objetivo claro de crear nuevas oportunidades de crecimiento y rentabilidad. (Kim & Mauborgne, 2005)

Los modelos de negocio innovadores están transformando la manera de hacer negocios en el mundo. Ante este nuevo escenario para los negocios, es necesario tener la capacidad de responder rápidamente a los requerimientos del mercado, y lograr la personalización de sus productos y servicios. (Osterwalder & Pigneur, 2011)

El conocer a sus clientes en tiempos de abundancia de datos implica tener herramientas de apoyo a la toma de decisiones, que les permita extraer y procesar grandes volúmenes de datos no estructurados, los cuales se encuentran disponibles en las plataformas de redes sociales a través de comentarios, publicaciones, imágenes, revisión de productos y videos publicados por los usuarios.

#### **1.1.5. Big Data**

La tecnología nos permite ahora capturar y almacenar grandes cantidades de datos. Encontrar patrones, tendencias y anomalías en estos conjuntos de datos, resumirlos con modelos cuantitativos simples es uno de los grandes desafíos de la era de la información: convertir los datos en información y convertir la información en conocimiento. (Witten & Frank, Data Mining Practical Machine Learning Tools and Techniques Second Edition, 2005)

La accesibilidad de los datos abre escenarios y oportunidades prometedores, y plantea una pregunta tentadora: ¿es posible convertir esos datos en información y conocimiento que luego puedan ser utilizados por los tomadores de decisiones para ayudar y mejorar la gobernanza de las empresas y de la administración pública? (Vercellis, 2009)

Lo que hace que los sistemas de información gerencial sean el tema más interesante en los negocios es el cambio continuo en la tecnología, el uso gerencial de la tecnología y el impacto en el éxito empresarial

Hay tres cambios interrelacionados en el área de tecnología: (1) la plataforma digital móvil emergente, (2) el creciente uso comercial de la "Big Data" y (3) el crecimiento de la "computación en la nube", donde cada vez hay más software de negocios corriendo a través de Internet. (Laudon & Laudon, 2014)

En el ámbito digital, los datos no estructurados están dominando los datos estructurados con un volumen de hasta un 80% en comparación con su contraparte (datos estructurados) con solo un 20%. Las redes sociales están ganando una popularidad masiva como medios de difusión de información y, además, como facilitadores de interacciones sociales. (Salloum, Al-Emran, & Shaalan, 2017)

La convergencia de la informática y la comunicación ha producido una sociedad que se alimenta de información. Sin embargo, la mayor parte de la información está en su forma cruda: datos. Si los datos se caracterizan como hechos registrados, entonces la información es el conjunto de patrones o expectativas que subyacen a los datos. Hay una gran cantidad de información encerrada en bases de datos que es potencialmente importante pero que aún no ha sido descubierta o articulada. (Witten & Frank, Data Mining Practical Machine Learning Tools and Techniques Second Edition, 2005)

La revisión de productos son una fuente importante de datos no estructurados, externa a los sistemas de soporte a las operaciones de

negocio, según (Pedrycz & Chen, 2016), la captura de los sentimientos y estados emocionales presentes en la información textual abarca una amplia gama de actividades orientadas a la web, como el análisis de sentimiento asociados a la revisión de productos, el desarrollo programas de marketing para mejorar los servicios al cliente.

Las revisiones de clientes en línea se definen como evaluaciones de productos o servicios realizadas por los clientes en sitios web y redes sociales. Las revisiones en línea están orientadas al usuario y describen el producto en varios escenarios de uso, lo evalúan desde la perspectiva del usuario. Actúan como recomendaciones y ayudan a los consumidores en su decisión de compra. (Manavi & Malini, 2018)

#### **1.1.6. Índice de promotores neto (NPS)**

El entorno actual se caracteriza por una sobreoferta de productos y servicios. Cada día es más costoso para las empresas poder encontrar y aún más difícil conservar a los clientes. Son el recurso más escaso y, en muchos sectores económicos, su captación se está convirtiendo en verdaderos juegos de suma cero, donde el total de clientes captados es igual al de los clientes perdidos. (Arjonilla, 2009)

El cliente ya no vive de forma pasiva, esperando recibir las ofertas; ha sabido adaptarse a un medio altamente competitivo y participa activamente, pero también se informa, compara, opina y recomienda. Al mismo tiempo apoyados por incremento del uso de Internet y la aparición de nuevas

tecnologías como las redes sociales la cual facilita y promueve las nuevas capacidades que caracterizan al consumidor de este siglo. (Arjonilla, 2009)

Existe un amplio consenso sobre la importancia de medir el concepto de compromiso de los clientes además de su satisfacción, y hacerlo con un indicador que combine satisfacción, fidelidad y predisposición a recomendar es el modelo más extendido. (Arjonilla, 2009)

La metodología más utilizada para medir la lealtad de los clientes es el Índice de Promotores Netos (NPS) por sus siglas en inglés, esta metodología fue desarrollada por Frederick Reichheld en el 2003. Esta metodología no se centra en la calidad, la satisfacción o el valor, sino se enfoca en medir como las recomendaciones de boca en boca del cliente tanto negativas como positivas pueden influenciar en el crecimiento de una empresa, servicio o producto. (Keiningham, Aksoy, Cooil, & Andreassen, 2008)

Los clientes califican su satisfacción en una escala de 0 a 10 sobre una empresa, un servicio o un producto que utilizan. Los resultados se utilizan para segmentar a los clientes en tres grupos: promotores, pasivos y detractores. (Jouve, Martin, & Guerin, 2012)

Los clientes que proporcionan una puntuación de 9 o 10 se consideran "promotores", 7 u 8 se consideran "pasivos " y 6 o menos "detractores". El indicador NPS se calcula restando el porcentaje de los promotores del porcentaje de los detractores. Por tanto, el NPS de una unidad o empresa

puede oscilar entre -100% y 100%. Además de la calificación en la escala de 10 puntos y el cálculo del NPS, es importante hacer una pregunta de seguimiento abierta: "¿Por qué proporcionó la calificación que proporcionó?" Además, como una extensión adicional, también es útil hacer una segunda pregunta abierta: "¿Qué se necesitaría para aumentar nuestra puntuación solo en un punto?" (Aguinis & Burgi-Tian, 2021)

Una de las características más atractivas de NPS es su simplicidad. Las empresas pueden eliminar la necesidad de realizar encuestas largas e incluso reducir sus presupuestos de investigación. La idea central de que cuantos más promotores tengas, más brillante será tu futuro. Pero la simplicidad no explica todo el atractivo de NPS. Lo que parece llamar más la atención de los gerentes corporativos es la afirmación de Reichheld sobre una relación entre NPS y crecimiento. (Keiningham, Aksoy, Cooil, & Andreassen, 2008)

La lealtad es la voluntad de alguien, un cliente, un empleado, un amigo, de hacer una inversión o un sacrificio personal para fortalecer una relación. Para un cliente, eso puede significar quedarse con un proveedor que lo trata bien y le brinda un buen valor a largo plazo, incluso si el proveedor no ofrece el mejor precio en una transacción en particular. (Reichheld, 2003)

La verdadera lealtad afecta claramente la rentabilidad. Si bien los clientes habituales no siempre son rentables, su elección de seguir con un producto o servicio generalmente reduce los costos de adquisición de nuevos clientes

de una empresa. La lealtad también impulsa el crecimiento de los ingresos brutos. (Reichheld, 2003)

## **1.2. DESCRIPCIÓN DE LA REALIDAD PROBLEMÁTICA**

Pocas organizaciones como las empresas de seguros tienen tanta información sobre sus clientes y consumidores y hacen tan poco con ella para generar una relación más cercana.

El contacto con el cliente está casi limitado a la publicidad, la web básicamente para presentar sus productos, el call center que los contacta para generar una venta, el agente comercial. Hay muchos otros momentos en los cuáles las empresas de seguros podrían estar en contacto con sus clientes y generar una relación más cercana. (Diario Gestión, 2017).

En este escenario de grandes desafíos que enfrentan ahora las organizaciones, se ha logrado identificar un conjunto de problemas o dificultades que no permiten a los directivos de las empresas a tomar decisiones en forma confiable, oportuna y basada en datos en sus diferentes niveles de toma de decisiones, los cuales se detallan a continuación:

- **Baja penetración del uso de herramientas de Inteligencia de Negocios en el Perú**

De acuerdo con la encuesta desarrollada por la consultora empresarial Ciclus Group a un grupo de 38 empresas peruanas cuyas facturaciones anuales fluctúan desde los 100 millones hasta los 5,000 millones de dólares,

la cual arrojó que en el Perú existe una baja penetración en el uso de herramientas de Inteligencia de Negocios.

El 52% de las empresas peruanas encuestadas, considera que el uso de herramientas de Inteligencia de Negocios se encuentra en un nivel básico, otro 42% estima que su uso es de nivel intermedio y solo un 5% se encuentra en una etapa avanzada. (Gestion.pe, 2013)

Por esta razón se hace necesario desarrollar herramientas de Inteligencia de Negocios que ayuden a los directivos de empresas a intensificar el uso de herramientas tecnológicas para la toma de decisiones, así como a recortar la brecha tecnológica que existe entre las empresas grandes y pequeñas.

- **Incremento del volumen de los datos no estructurados (Big Data)**

Ante el creciente dominio del volumen de los datos no estructurada (Big Data), según los autores (Salloum, Al-Emran, & Shaalan, 2017) los datos no estructurados representan el 80% en comparación al 20% de los datos estructurados.

Debido a este incremento en el volumen de datos no estructurados se hace casi imposible para las empresas poder procesar y obtener información de valor para la toma de decisiones, usando bases de datos, herramientas y programas de computador de manera tradicional.

Los datos no estructurados son también producidos diariamente por la interacción de los usuarios en redes sociales, en los últimos años las redes

sociales han adquirido un nivel de importancia en la vida de las personas, las personas se sienten cómodas de publicar comentarios, actualizar estados y subir fotos y videos.

Además, los usuarios demuestran su sentimiento de manera positiva, negativa o neutra reaccionando ante un comentario o publicación debido a un cambio en la coyuntura política, social o económica o por alguna queja, denuncia o problema presentado por una marca o empresa.

La conectividad y la facilidad de acceso a Internet están originando que la cantidad de usuarios que se conectan e interactúan en las redes sociales siga incrementándose mes a mes.

Como muestran las cifras estadísticas de Facebook que hasta al 31 de diciembre del 2020, existen más de 2,797 billones de usuarios activos que se conectan mensualmente (Monthly Active Users – MAU por sus siglas en inglés).

Esto ha significado un incremento del 11% en el MAU de Facebook por cada año, comparado con el cuarto trimestre del 2019 que fue de 2,498 mil millones de MAU (Facebook.com, Facebook Newsroom, 2019) (Statista.com, 2021)

En el Perú, existen 24 millones de usuarios activos en redes sociales lo cual representa un 73% de la población. 23 millones de estos usuarios activos

utilizan un teléfono inteligente para acceder a internet, lo cual representa un 70% de la población.

Adicionalmente en el Perú existen 39.08 millones de dispositivos móviles, lo cual significa que existen más celulares que habitantes, esto representa un 119% como porcentaje de la población (wearesocial.com & hootsuite.com, 2019)

Adicionalmente, Facebook es una de las plataformas de redes sociales que tiene una gran cantidad de interacción de sus usuarios, cada 60 segundos se publican 510,000 comentarios, se actualizan 293,000 estados y se suben 136,000 fotos. (Zephoria.com, The Top 20 Valuable Facebook Statistics – Updated May 2019, 2019)

- **No se tiene reacción inmediata ante campañas negativas en redes sociales**

Este nuevo comportamiento que tiene el consumidor en redes sociales ha originado un cambio en la forma como las empresas deben hacer negocios. Las empresas que tienen presencia en redes sociales deben conocer inmediatamente los comentarios de los usuarios en redes sociales con respecto a una marca, producto o servicio para poder reducir el tiempo de respuesta a esos comentarios.

Conocer los comentarios negativos de manera oportuna permitirá diseñar una estrategia para enfrentar las amenazas de una campaña de

desprestigio. Responder y atender rápidamente sus requerimientos permitirá reducir los efectos negativos de una campaña.

Este conocimiento también permitirá eliminar o reducir ciertos atributos de un producto o servicio que no son percibidos de valor por el consumidor, para incrementar o adicionar los atributos que están por debajo de la industria de empresas corredoras de seguros o para diseñar nuevos productos y servicios que cubran los nuevos requerimientos de los consumidores.

- **No se cuenta con capacidad de procesamiento avanzado para transformar comentarios subjetivos en información de valor para la toma de decisiones**

Miles de millones de personas utilizan las redes sociales diariamente para expresar sus emociones, opiniones y revelar aspectos comunes de su vida cotidiana de una manera natural, las personas acostumbran a publicar o realizar comentarios sobre productos y servicios a través de las comunidades en línea, en donde los consumidores informan e influyen a otros usuarios. (Sarlan, Nadam, & Basri, 2014)

Los comentarios de los usuarios son de naturaleza subjetiva y están inherentemente cargados de emociones y sentimiento. También pueden ser contradictorios e inconsistentes. Dichos comentarios se presentan en la web y se necesitarán herramientas de procesamiento avanzadas y de análisis de sentimiento para que podamos transformarlos en puntos de vista coherentes y significativos acerca de productos y servicios (Pedrycz & Chen, 2016).

- **Transformación digital en las empresas del sector asegurador**

El cambio en el comportamiento en el consumidor, las altas expectativas que tienen en relación con los servicios que contratan y la posibilidad que ahora poseen de hacerse escuchar a gran escala, son un reto para las empresas de seguros, pero son a la vez una gran oportunidad.

Everis presenta estas oportunidades para transformar a las empresas de seguros en tres puntos:

**Redes Sociales:** La transformación de la experiencia del cliente, la conversación social para el diseño de nuevos productos y la creación de una fuerte diferenciación usando atributos emotivos.

**Omnicanalidad:** El consumidor opera de manera natural a través de múltiples dispositivos que alterna de forma fluida a lo largo del día en función del contexto y de sus necesidades. Las empresas del sector asegurador deben garantizar que cuando el usuario entre a través de su móvil encuentre información de la compañía, productos, los prestadores del servicio, etc. La cotización, emisión y pago, y la gestión de siniestros.

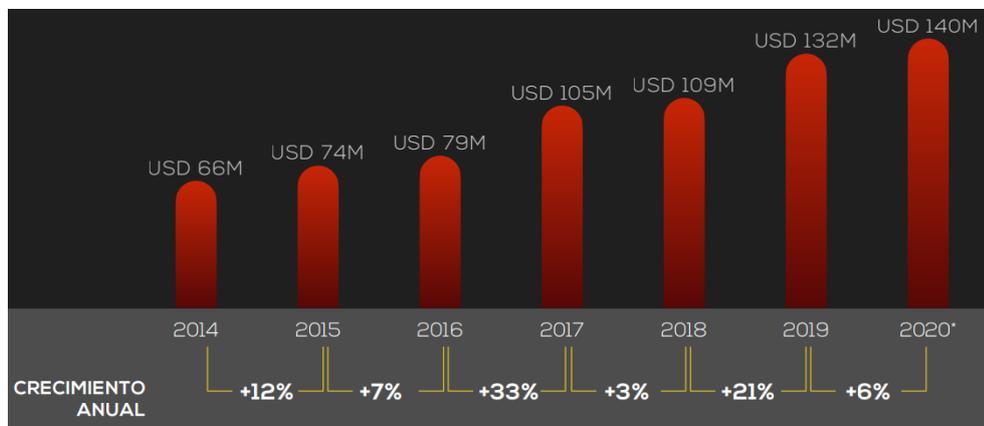
**Analytics:** Las aseguradoras manejan enormes volúmenes de datos, fruto de las interacciones y transacciones que se registran automáticamente en servidores. La incorporación de capacidades analíticas avanzadas en la base de las aseguradoras dará ventajas competitivas, ya que establece una base robusta sobre la que construir estrategias de fidelización, retención y crecimiento.

### 1.2.1. Inversión en publicidad digital

- **Evolución de la Inversión**

La inversión publicitaria digital en el Perú creció 6% en el 2020 pese al contexto adverso que enfrentamos en el país y el mundo, como es la pandemia del coronavirus. De acuerdo a la Figura N° 1 el crecimiento en inversión publicitaria fue de USD 8MM.

Figura N° 1: Evolución de la Inversión Anual, 2014 - 2020 (En Millones de dólares)

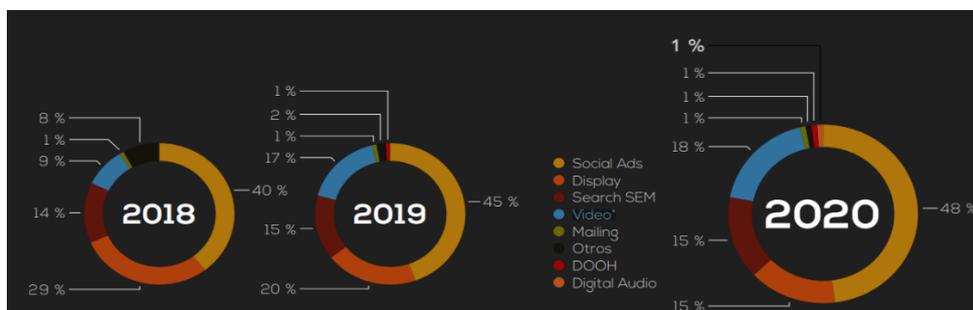


Fuente: IAB Perú - Informe de Inversión Publicitaria 2020

- **Comparativo de formatos**

Como se puede apreciar en la Figura N° 2, la inversión en publicidad digital tuvo sorpresas en el 2020, se destaca el crecimiento del formato Video con (+6%) y el formato Social Ads con (+7%), mientras que el formato Display decreció un (-25%) que obtuvo una participación de 15%. Vale destacar que ambos formatos representan más del 63% de la inversión.

Figura N° 2: Comparativo de Formatos, 2018 - 2020 (Porcentaje)



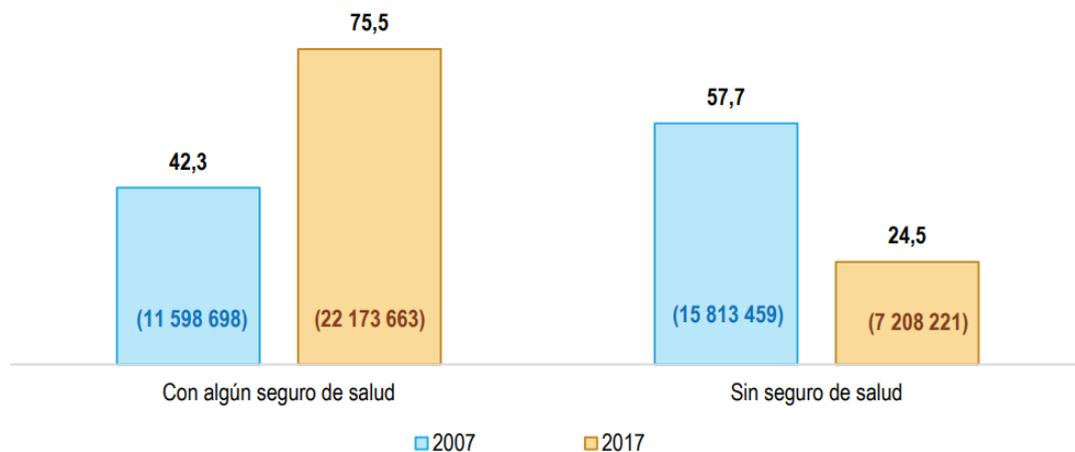
Fuente: IAB Perú - Informe de Inversión Publicitaria 2020

### 1.2.2. Tenencia de seguro de salud

De acuerdo con los resultados del censo 2017, del total de la población censada, el 75,5% respondió tener algún seguro de salud y el 24,5% que no tenía, es decir, no está asegurada.

En el período intercensal 2007-2017, como se aprecia en la figura N° 3, hay una disminución de 8,605,238 personas sin seguro, al pasar de 15,813,459 (57,7%) a 7,208,221 (24,5%) que representa 33,2 puntos porcentuales.

Figura N° 3: Población Censada Según Condición de Tenencia de Seguros de Salud, 2007 y 2017 (Absoluto y Porcentaje)



Fuente: INEI - Censos Nacionales de Población y Vivienda 2007 y 2017.

- **Grupos de edad**

Los grupos de población menor de 15 años de edad y los adultos mayores de 70 y más años son los que muestran menor proporción de la población sin seguro (por debajo del 20,0%), según el Censo del 2017.

Por otro lado, según se aprecia en la Figura N° 4, con más del 30,0% de personas sin seguro, se encuentran en los grupos de 20 a 24 y de 25 a 29 años.

Figura N° 4: Población Censada sin Seguro de Salud, Según Grupos de Edad, 2017 (Absoluto y porcentaje)

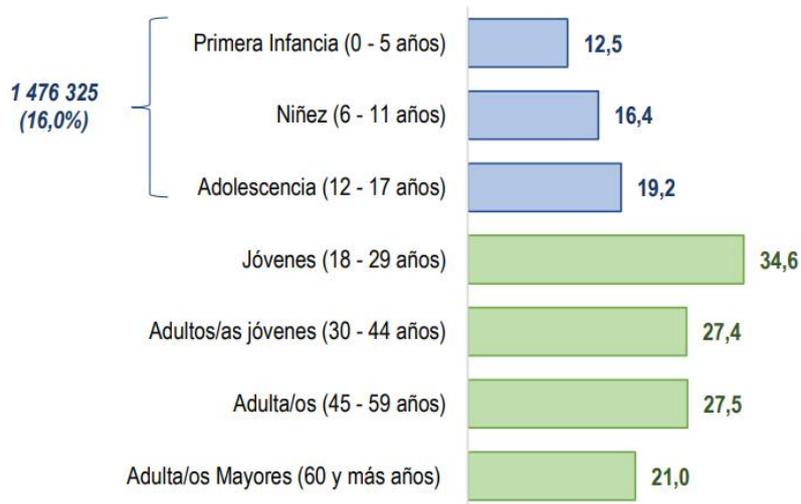


Fuente: INEI - Censos Nacionales 2017: XII de Población y VII de Vivienda.

- **Ciclo de vida**

Como se aprecia en la Figura N° 5, la población de niños y adolescentes sin seguro de salud totaliza 1,476,325 y representa el 16,0% del total de la población censada de este grupo. El 34,6% de los jóvenes de 18 a 29 años de edad no cuentan con seguro de salud.

Figura N° 5: Población Censada sin Seguro de Salud, Según Ciclo de Vida, 2017 (Porcentaje respecto al total de población censada por grupo de edad)



Fuente: INEI - Censos Nacionales 2017: XII de Población y VII de Vivienda.

- **Nivel educativo**

El Censo 2017 da cuenta que entre las personas de 15 y más años de edad que no tienen ningún nivel educativo, como se puede apreciar en la Figura N° 6, el 21,9% (237,345) no están aseguradas y entre las que registraron educación básica especial, este porcentaje alcanzó 32,7%.

Figura N° 6: Población Censada de 15 y más años, sin seguro de salud, según nivel educativo, 2017 (Absoluto y porcentaje)



Fuente: INEI - Censos Nacionales 2017: XII de Población y VII de Vivienda.

### 1.2.3. Población ocupada con seguros de salud

En el año 2017, se registró 11,985,400 personas, como se puede apreciar en la Figura N° 7, representan el 72.6% de la PEA, los cuales tiene alguno de los siguientes seguros de salud:

- Seguro Social de Salud (EsSalud)
- Seguro Integral de Salud (SIS)
- Otro seguro:
  - Seguro Privado de Salud
  - Entidad Prestadora de Salud
  - Seguro de las Fuerzas Armadas/Policiales
  - Seguro Universitario
  - Seguro Escolar Privado
  - Otros

Mientras que el 27.4% carece de este beneficio.

Figura N° 7: Población ocupada con seguro de salud, según área de residencia, 2007, 2016, 2017 (Miles de personas y porcentajes)

Área de residencia	2007	2016	2017	Tasa de crecimiento promedio anual (%) 2007-2017	Variación porcentual (%) 2017/2016
<b>Total</b>	<b>4 299,0</b>	<b>11 700,0</b>	<b>11 985,4</b>	<b>10,8</b>	<b>2,4</b>
(% del total de ocupados)	30,3	72,2	72,6		
<b>Área de residencia</b>					
Urbana	<b>3 535,6</b>	<b>8 675,1</b>	<b>8 942,2</b>	<b>9,7</b>	<b>3,1</b>
(%)	35,3	70,0	70,3		
Rural	<b>763,5</b>	<b>3 024,9</b>	<b>3 043,3</b>	<b>14,8</b>	<b>0,6</b>
(%)	18,2	79,4	80,3		

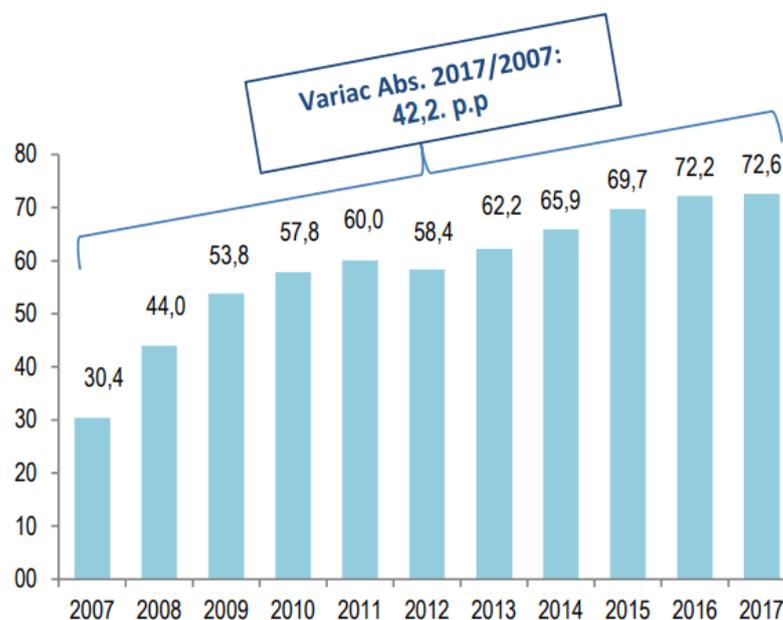
Fuente: Instituto Nacional de Estadística e Informática-Encuesta Nacional de Hogares.

- **Tasa de crecimiento del Seguro Integral de Salud (SIS) es mayor al del Seguro Social de Salud (EsSalud).**

Según cifras del INEI, en el año 2017, se registró 11'985,400 personas, que representan el 72.6% de la población ocupada, tienen algún seguro de salud, mientras que el 27.4% carece de este beneficio.

En el año 2017, el 72.6% de los ocupados cuentan con algún seguro de salud, el que comparado al registrado en el 2007 (30.4%) representa un incremento de 42.2 puntos porcentuales, como se puede apreciar la Figura N° 8.

Figura N° 8: Población ocupada con seguro de salud, 2007-2017 (porcentaje del total de ocupados)



Fuente: Instituto Nacional de Estadística e Informática-Encuesta Nacional de Hogares.

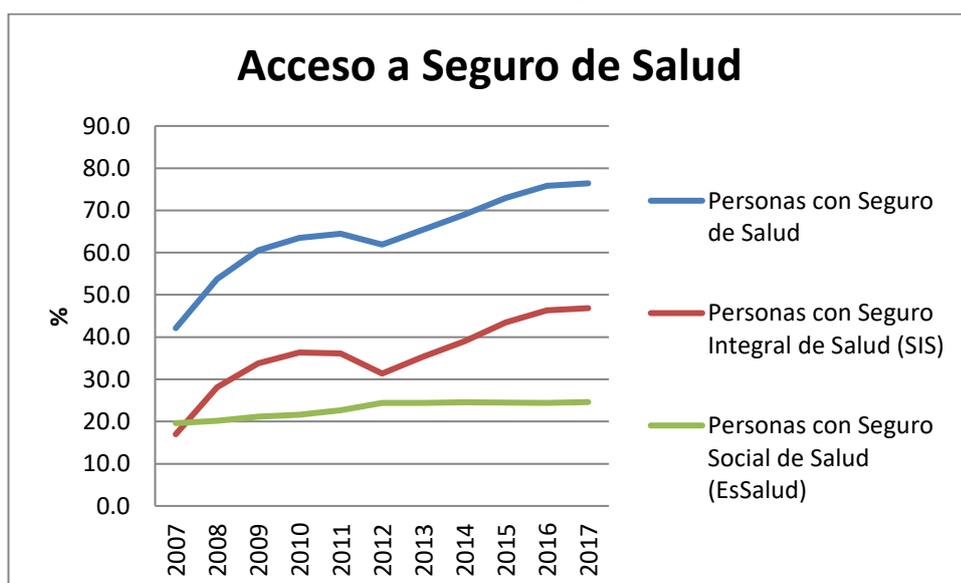
Sin embargo, se puede visualizar que el incremento de la población ocupada con algún seguro de salud se debe al incremento de la cobertura del Seguro Integral de Salud (SIS), cuya tasa de crecimiento promedio, para el periodo

analizado, fue de 24.0%, como se puede apreciar en la Figura N° 9, está muy por encima de la tasa de crecimiento del Seguro Social de Salud (EsSalud), 4.3%.

Este incremento de dos cifras en el seguro SIS podría provocar un colapso del seguro integral de salud SIS, por la falta de infraestructura en los hospitales (Buscar crecimiento de número de camas en los hospitales) y falta de medicina, debido a la deuda que mantiene el estado con la asociación de proveedores de medicina (Buscar deuda del estado con los proveedores del SIS).

El propósito de la creación del seguro de SIS era atender a las personas en situación de pobreza extrema, quienes no tiene recursos para cotizar un seguro de salud.

Figura N° 9: Acceso a Seguro de Salud



Fuente: Instituto Nacional de Estadística e Informática  
Elaboración Propia

### 1.2.4. Incremento del empleo informal

En el año 2017, según resultados de la Encuesta Nacional de Hogares (ENAH), como se puede apreciar en la Figura N° 10, la población ocupada en empleo informal alcanzó los 11,978,100 personas, que representan el 72.5% del total de ocupados y 4,532,800 personas, que equivale al 27.5%, se encuentran laborando en empleos formales de la economía.

Figura N° 10: Población ocupada por empleo formal e informal, según ámbito geográfico 2008, 2016 y 2017 (miles de personas)

Ámbito geográfico	2008			2016			2017			Tasa de crecimiento promedio anual 2007-2017 (%)		Variación % 2017/2016	
	Total	Empleo formal	Empleo informal	Total	Empleo formal	Empleo informal	Total	Empleo formal	Empleo informal	Empleo formal	Empleo informal	Empleo formal	Empleo informal
<b>Total</b>	14 459,6	3 012,7	11 446,9	16 197,1	4 539,8	11 657,3	16 511,0	4 532,8	11 978,1	4,6	0,5	-0,2	2,8
<b>Área de residencia</b>													
Urbana	10 303,6	2 869,1	7 434,5	12 389,5	4 360,5	8 029,1	12 721,5	4 362,4	8 359,0	4,8	1,3	0,1	4,1
Rural	4 156,0	143,6	4 012,4	3 807,6	179,3	3 628,2	3 789,5	170,4	3 619,1	1,9	-1,1	-5,0	-0,3
<b>Región natural</b>													
Costa	7 651,3	2 153,0	5 498,3	8 841,9	3 344,6	5 497,4	9 027,3	3 327,7	5 699,6	5,0	0,4	-0,5	3,7
Sierra	4 917,7	625,3	4 292,4	5 253,5	871,9	4 381,6	5 337,9	875,0	4 462,9	3,8	0,4	0,4	1,9
Selva	1 890,6	234,4	1 656,2	2 101,7	323,3	1 778,4	2 145,8	330,2	1 815,6	3,9	1,0	2,1	2,1

Fuente: Instituto Nacional de Estadística e Informática - Encuesta Nacional de Hogares.

### 1.2.5. Baja profundidad de la inclusión financiera

- **Acceso al sistema financiero no formal**

La información de la Encuesta Nacional de Hogares revela que los trabajadores independientes hacen uso del financiamiento no formal para el desarrollo de su actividad emprendedora, una fuente importante de financiamiento son las que provienen de las denominadas:

- Juntas
- Ahorros propios guardado en casa

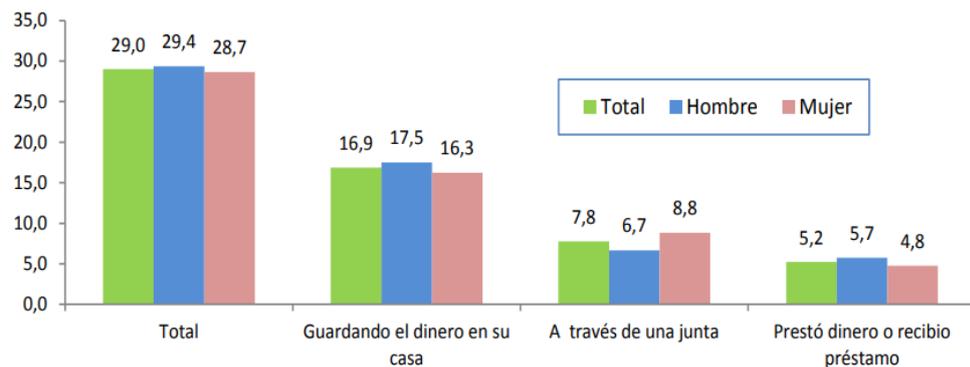
- Si prestó dinero o recibió préstamo
- Guardó dinero con familiares

Los resultados del año 2017, como se puede apreciar en la Figura N° 11, el 29.0% de los trabajadores independientes hacen uso de recursos no formales de financiamiento.

Entre las formas de obtener recursos financieros se muestra que el 16.9% guardó dinero en su casa, el 7.8% obtuvo recursos a través de una “Junta” y el 5.2% prestó dinero o recibió préstamo.

Cabe resaltar que en relación con las “juntas”, son las mujeres (8.8%) en comparación a los hombres (6.7%) las que hacen un mayor uso de esta forma de obtener recursos.

Figura N° 11: Población ocupada en condición de trabajador independiente por sexo, según acceso a otras formas de servicio financiero no formales, 2017 (Porcentaje)



Fuente: Instituto Nacional de Estadística e Informática-Encuesta Nacional de Hogares.

### 1.2.6. El futuro del seguro de automóvil conectado

Ante el aumento de automóviles y teléfonos inteligentes conectados a internet son uno de los desafíos que enfrentan las empresas de seguros,

ofrecer productos y servicios que se basen en la tecnología para que sus clientes puedan a reducir la cotización en el pago de sus primas.

Según el informe de Forbes, manejar de forma segura le puede ahorrar hasta un 30% en las primas de seguros. Esto es el futuro del seguro de automóvil conectado, también denominado seguro de automóvil basado en el uso, conocido también por sus siglas en inglés como: UBI, Usage-Based Insurance. (National Association Insurance Commissioners, 2019).

Los clientes podrán tener una cotización competitiva en función de cómo maneja, cuándo maneja y donde maneja. Se puede acceder a la información de conducción en línea o en aplicaciones que permiten a los clientes monitorear sus patrones de conducción y hacer los ajustes necesarios para mejorar sus posibilidades de obtener mejores descuentos. (Forbes.com, 2017)

### **1.3. FORMULACIÓN DEL PROBLEMA**

En el ámbito del diseño de productos y servicios, son varias empresas que recurren a sociólogos expertos para comprender mejor el panorama social. Este enfoque puede propiciar la aparición de modelos de negocios nuevos o mejores. (Osterwalder & Pigneur, 2011)

Desde su aparición, las redes sociales fueron utilizadas solo como un lugar donde las personas se sientan cómodas de participar, compartir y crear contenido.

Ahora las redes sociales también son usadas por empresas como un medio para promover campañas de marketing y tener contacto cercano con sus consumidores a través una comunicación directa que ofrecen las redes sociales (Kurnia & Suharjito, 2018).

Los consumidores están dispuestos a compartir contenido de sus marcas debido a tres motivaciones: social, emocional y funcional. El motivador social se relaciona con el envío de señales a otros sobre la experiencia, la singularidad o el estatus social, el motivador emocional está relacionado con la necesidad de compartir sentimientos positivos o negativos sobre las marcas y el motivador funcional está relacionado con la tendencia de obtener y proporcionar información a otros. (Lovett, Peres, & Shachar, 2013)

Para diseñar productos y servicios o modificar los existentes, es de vital importancia conocer el grupo de clientes a quiénes estará dirigido. El mejor lugar donde encontrarlos e identificar su comportamiento, son las redes sociales.

Para eliminar y reducir atributos que no son percibidos de valor por el consumidor, para incrementar o crear los atributos que están por debajo de la industria de los seguros, para diseñar y lanzar al mercado nuevos e innovadores programas de seguro que resulten rentables para las empresas corredoras de seguro es necesario conocer el comportamiento que tienen los consumidores en redes sociales acerca de nuestra marca.

Uno de los desafíos que enfrentan las empresas de seguros es la diseñar y ofrecer servicios apoyados por la tecnología para que sus clientes puedan a reducir la cotización en el pago de sus primas de seguro. Según el informe de Forbes, manejar de forma segura le puede ahorrar hasta un 30% en las primas de seguros.

El futuro del seguro de automóvil conectado, también se le conoce como seguro de automóvil basado en el uso (UBI Usage-Based Insurance) por sus siglas en inglés. Los clientes podrán tener una cotización más competitiva en función de cómo maneja, cuándo y adonde maneja.

#### **1.3.1. Problema general**

¿Contribuye un modelo de inteligencia de negocios basado en minería de opinión sobre redes sociales, a la mejora del servicio al cliente en una empresa corredora de seguros en el Perú?

#### **1.3.2. Problema específico**

PE1: Los comentarios se encuentran dispersos en las diferentes cuentas que las empresas corredoras de seguros tienen en redes sociales.

PE2: Los comentarios está llenos de opiniones son de naturaleza subjetiva, contradictorios e inconsistentes, lo cual no permite su clasificación en una categoría o clase.

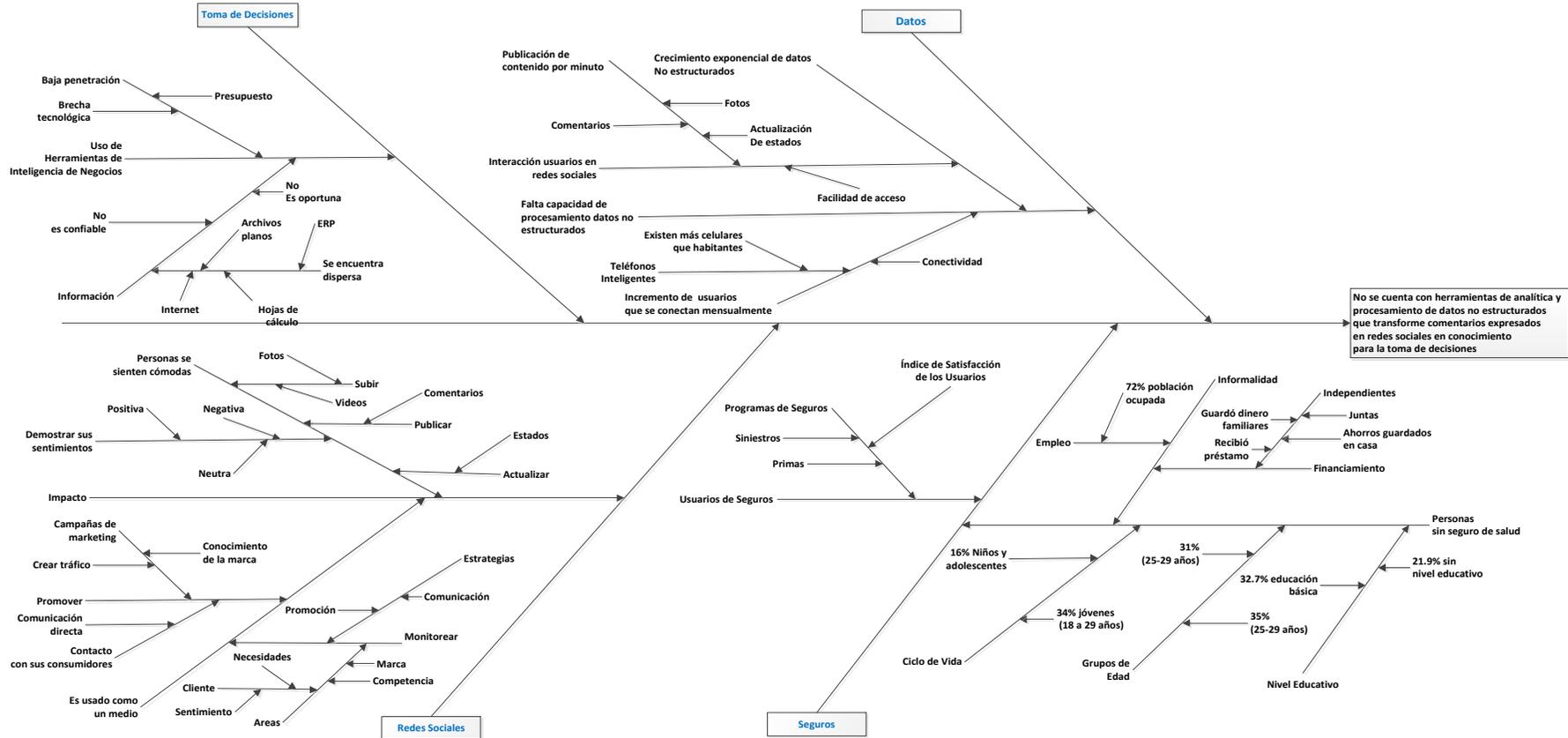
PE3: Los comentarios están llenos de sentimientos, que es una emoción subjetiva, lo cual no permite identificar las emociones que expresan los usuarios en sus comentarios en redes sociales.

PE4: No se cuenta con indicadores que permitan realizar minería de opinión sobre comentarios expresados en redes sociales que contribuya a la mejora del servicio al cliente en una empresa corredora de seguros.

### 1.3.3. Diagrama de Ishikawa

Como se puede visualizar en la Figura N° 12, se ha realizado el estudio de la problemática del sector asegurador en nuestro país utilizando el diagrama de Ishikawa como herramienta para identificar el problema principal y las posibles causas que las generan.

Figura N° 12: Diagrama de Ishikawa



Fuente: Elaboración Propia

#### **1.4. DELIMITACIÓN DE LA INVESTIGACIÓN**

La presente investigación está basada en desarrollar un modelo de minería de opinión sobre comentarios expresados en la plataforma de red social: Facebook, acerca de los servicios prestados por las empresas corredoras de seguros en Lima Perú.

#### **1.5. JUSTIFICACIÓN E IMPORTANCIA DE LA INVESTIGACIÓN**

La importancia de esta investigación radica en hacer inteligencia de negocios social que ayude a las empresas del sector asegurador a identificar, conocer y segmentar a sus clientes potenciales, poder entender sus motivaciones, necesidades y requerimientos, que les permita crear una propuesta de valor acorde a sus necesidades.

Esta investigación ayudará a los clientes potenciales en redes sociales a conocer más acerca de los diversos productos y servicios del sector asegurador y la importancia que tienen para los consumidores y sus familias en su vida diaria el adquirir un seguro.

Las empresas del sector asegurador podrán comunicarse directamente con sus clientes a través de campañas publicitarias, la creación de contenido en diferentes plataformas de redes sociales el cual estará dirigido a un segmento de mercado específico lo cual mejorará el posicionamiento de la marca y mejorará la efectividad de las campañas de marketing digital.

Uno de los desafíos a los que se enfrentan las empresas del sector asegurador es diseñar productos y servicios apoyados por la tecnología para que sus clientes puedan a reducir el pago de sus primas.

## **1.6. OBJETIVOS DE LA INVESTIGACIÓN**

### **1.6.1. Objetivo general**

Desarrollar un Modelo de Inteligencia de Negocios basado en Minería de Opinión sobre redes sociales para la mejora del servicio al cliente, en una empresa corredora de seguros en el Perú.

### **1.6.2. Objetivos específicos**

OE1: Extraer comentarios desde las diferentes cuentas que las empresas corredoras de seguros tienen en redes sociales para integrar las fuentes de en un repositorio de datos.

OE2: Seleccionar el algoritmo de clasificación que tenga el mejor valor de precisión para clasificar un comentario en una categoría.

OE3: Construir un modelo de análisis de sentimiento que transforme comentarios en información objetiva para identificar emociones que expresan los usuarios en redes sociales.

OE4: Diseñar un tablero de control que muestre indicadores de opiniones y emociones expresadas en redes sociales para contribuir a la mejora del servicio al cliente en una empresa corredora de seguros.

## **1.7. HIPÓTESIS DE LA INVESTIGACIÓN**

### **1.7.1. Hipótesis general**

Un Modelo de Inteligencia de Negocios basado en minería de opinión sobre redes sociales contribuirá a la mejora del servicio al cliente en una empresa corredora de seguros en el Perú.

### **1.7.2. Hipótesis específicas**

H1: La extracción de comentarios desde las diferentes cuentas que las empresas corredoras de seguros tienen en redes sociales permitirá integrar las diferentes fuentes en un repositorio de datos.

H2: El algoritmo de clasificación de texto que tenga mejor valor de precisión permitirá la clasificación de un comentario en una categoría o clase.

H3: Un modelo de análisis de sentimiento que transforme comentarios en sentimiento permitirá identificar las emociones que expresan los usuarios en redes sociales.

H4: El diseño de un tablero de control que muestra los indicadores de opiniones y emociones expresadas en redes sociales contribuye a la mejora del servicio al cliente en una empresa corredora de seguros.

## **1.8. VARIABLES DE INDICADORES**

### **1.8.1. Identificación de las variables**

**Variable independiente X:** Modelo de Inteligencia de Negocios

**Variable dependiente Y:** Contribuye a la mejora del Servicio al Cliente

### 1.8.2. Operacionalización de variables

Para descomponer deductivamente las variables que componen el problema de la investigación, se ha utilizado el proceso metodológico de operacionalización de variables Tabla N° 1.

Tabla N° 1: Matriz de operacionalización de variables

VARIABLES	DIMENSIONES	INDICADORES	UNIDADES/ CATEGORÍAS	TIPO	OBJETIVO
X: Modelo de Inteligencia de Negocios	Fuente de Datos	%Integración	Porcentaje	Numérica continua	O1
	Comentarios	%Precisión	Porcentaje	Numérica continua	O2
			Porcentaje	Numérica continua	
		%Precisión	Positivo Negativo Neutro	Numérica discreta Numérica discreta Numérica discreta	O3
Y: Contribuye a la mejora del Servicio al Cliente	Servicio al Cliente	Servicio al Cliente	Porcentaje	Numérica continua	O4

Fuente: Elaboración Propia (2020)

### 1.8.3. Matriz de consistencia

En la Tabla N° 2 se podrá encontrar la relación lógica que existe entre la definición del problema, los objetivos, las hipótesis y las variables dependientes e independientes de la presente investigación.

Tabla N° 2: Matriz de consistencia

<b>PROBLEMA</b>	<b>OBJETIVO</b>	<b>HIPÓTESIS</b>	<b>VARIABLES</b>
¿Contribuye un Modelo de Inteligencia de Negocios basado en Minería de Opinión sobre redes sociales, a la mejora del servicio al cliente, en una empresa corredora de seguros en el Perú?	Desarrollar un modelo de Inteligencia de Negocios basado en Minería de Opinión sobre redes sociales para la mejora del servicio al cliente, en una empresa corredora de seguros en el Perú.	Un modelo de Inteligencia de Negocios basado en Minería de Opinión sobre redes sociales contribuye a la mejora del servicio al cliente, en una empresa corredora de seguros en el Perú.	<b>Independiente:</b> Modelo de Inteligencia de Negocios  <b>Dependiente:</b> Contribuye a la mejora del Servicio al Cliente.
PE1: Los comentarios se encuentran dispersos en las diferentes cuentas que las empresas corredoras de seguros tienen en redes sociales.	OE1: Extraer comentarios desde las diferentes cuentas que las empresas corredoras de seguros tienen en redes sociales para integrar las fuentes en un repositorio de datos.	HE1: La extracción de comentarios desde las diferentes cuentas que las empresas corredoras de seguros tienen en redes sociales permitirá integrar las diferentes fuentes en un repositorio de datos.	%Integración
PE2: Los comentarios están llenos de opiniones que son de naturaleza subjetiva, contradictorios e inconsistentes, lo cual no permite su clasificación en	OE2: Seleccionar el algoritmo de clasificación de texto que tenga el mejor valor exactitud y precisión para clasificar un comentario en una categoría o clase.	HE2: El algoritmo de clasificación de texto que tenga mejor valor de precisión permitirá la clasificación de un comentario en una categoría o clase.	%Precisión

una categoría o clase.			
PE3: Los comentarios están llenos de sentimiento que es una emoción subjetiva, lo cual no permite identificar las emociones que expresan los usuarios en sus comentarios en redes sociales.	OE3: Utilizar técnicas de análisis de sentimiento que transforme comentarios en información objetiva para identificar las emociones que expresen los usuarios en redes sociales.	HE3: El uso de técnicas de análisis de sentimiento que transforme comentarios en sentimiento, permitirá identificar las emociones que expresan los usuarios en redes sociales.	Sentimientos: Positivo, Negativo, Neutro
PE4: No se cuenta con indicadores que permitan realizar minería de opinión sobre comentarios expresados en redes sociales que contribuya a la mejora del servicio al cliente en una empresa corredora de seguros.	OE4 Diseñar un tablero de control que muestre indicadores de opiniones y emociones expresadas en redes sociales para contribuir a la mejora del servicio al cliente en una empresa corredora de seguros.	HE4: El diseño de un tablero de control que muestre los indicadores de opiniones y emociones expresadas en redes sociales permitirá contribuye a la mejora del servicio al cliente en una empresa corredora de seguros.	Servicio al Cliente

Fuente: Elaboración Propia (2020)

## **CAPÍTULO II**

### **MARCO DE REFERENCIA**

#### **2.1. ANTECEDENTES DE LA INVESTIGACIÓN**

##### **Antecedente N°1, Kurnia, P. F., & Suharjito. (2018)**

Título: Business Intelligence Model to Analyze Social Media Information

Tipo de Documento: Artículo Procedia Computer Science 135 (2018) 5–14

Objetivo: <<Analizar el contenido de las redes sociales usando minería de texto para obtener información de las redes sociales para la toma de decisiones que puede traer beneficios a las personas y a las empresas>>

Metodología: Investigación Descriptiva

Resultados: En la presente investigación, la evaluación que se usó para el algoritmo de clasificación es el esquema de prueba con 5 composiciones de datos diferentes.

Para cada prueba, la composición se realizó aplicando un método de validación cruzada de 10 veces en los algoritmos Naive Bayes, Decision Tree y SVM (Support Vector Machine).

El algoritmo de prueba de Naive Bayes con Cross Validation con los datos de la tercera prueba en la tercera iteración es la que tiene la mejor precisión, con un valor de precisión de 80.141%. Sin embargo, si se observa el promedio de todas las pruebas, se obtuvo el mejor resultado de precisión en los datos de la cuarta prueba con un valor de precisión de igual a 74.787%.

Relación con la Investigación: Lograr clasificar y categorizar con una alta precisión el contenido publicado por usuarios en las redes sociales, nos permitirá determinar el tipo de información y la categoría a la que pertenece un determinado dato, luego estos datos podrán ser usados mejorar o diseñar nuevos programas de seguros, así como para mejorar el nivel de servicio al cliente.

De acuerdo con las evaluaciones realizadas por los autores del presente artículo, el algoritmo de clasificación: Naive Bayes con Cross Validation es la que tiene una mejor precisión para la clasificación de texto, este algoritmo será utilizado para la clasificación de texto.

**Antecedente N°2, Muntean, M., Cabău, L., & Rînciog, V. (2014)**

Título: Social Business Intelligence: A New Perspective for Decision Makers

Tipo de Documento: Artículo Procedia - Social and Behavioral Sciences 124  
562 – 567

Objetivo: <<Desarrollar inteligencia social basada en información obtenida de redes sociales a través de la recopilación, clasificación e integración con datos internos y externos tradicionales >> (Muntean, Cabău, & Rînciog, 2014)

Metodología: Investigación Descriptiva

Resultados: La integración de datos sociales con datos internos y externos tradicionales establece una nueva dimensión que puede introducirse en el marco de toma de decisiones.

En la presente investigación se demuestra que las personas se sienten cómodas intercambiando datos y puntos de vista a través de entornos sociales. Las empresas y las personas pueden desarrollar "inteligencia social" basada en la información extraída de redes sociales

Se ha descubierto los beneficios de escuchar a las redes sociales, las herramientas adecuadas, los sistemas y las tecnologías que ayudan a "separar la señal del ruido".

La información sobre los clientes, sus intereses y las decisiones de compra podrían transformarse en conocimiento, sumarse a la base de conocimiento organizacional y las decisiones de dominio de tierra.

Las capacidades analíticas integradas, los indicadores clave de rendimiento (KPI) para evaluar objetivamente los datos se ponen a disposición de quienes toman las decisiones.

Relación con la Investigación: Escuchar a las redes sociales, categorizarlos y transformarlos para obtener conocimiento preciso del comportamiento del consumidor se relaciona directamente con la presente investigación.

### **Antecedente N°3, Nick Hajli, Michel Larocheb (2018)**

Título: Applications of business intelligence and analytics in social media marketing

Tipo de Documento: Artículo International Journal of Information Management

Objetivo: <<Utilizar las grandes cantidades de datos de las redes sociales para captar la percepción de los consumidores de la gran variedad de contenido generado por los usuarios en las plataformas de las redes sociales, y examinar cómo BI&A (Business Intelligence and Analytics) permite a las empresas crear valor comercial y mantener una ventaja competitiva>>. (Hajli & Laroche, 2019)

Metodología: Investigación Descriptiva

Resultados: La Inteligencia de Negocios y la Analítica son retos que tienen las empresas quienes buscan adoptar un enfoque reflexivo y holístico para lograr el análisis y el aprovechamiento de los datos de las redes sociales.

Se ha encontrado que existen varios obstáculos importantes, entre los que se incluyen la falta de integración de datos, problemas de sobrecarga de datos, barreras para la recopilación de datos de los consumidores de alta calidad, y la cultura organizacional y la gestión del cambio.

El valor que tienen los datos de las redes sociales muy pocas veces se logra descubrir, analizar y visualizar, ya sea para crear conocimientos que complementen la falta de información para la toma de decisiones.

Relación con la investigación: La Inteligencia de Negocios Social y la Analítica de los datos son herramientas que ayudarán a las empresas en la

creación de valor para los negocios y para mantener una ventaja competitiva sobre su competencia. Uno de los aportes de la presente investigación será la exploración de los datos de las redes sociales para descubrir y captar las percepciones de los consumidores sobre una marca o servicio. la cual servirá para la mejora del nivel de servicio al cliente.

**Antecedente N°4, Sarlan, A., Nadam, C., & Basri, S. (2014)**

Título: Twitter Sentiment Analysis 212-216. 10.1109/ICIMU.2014.7066632.

Tipo de Documento: Artículo Procedia Computer Science 135 (2018) 5–1

Objetivo: <<Realizar minería de opinión sobre textos a través del procesamiento del lenguaje, minería de texto, lingüística computacional, que implica el estudio computacional de los sentimientos, opiniones y emociones expresadas en párrafos de texto>> (Sarlan, Nadam, & Basri, 2014)

Metodología: Investigación Descriptiva

Resultados: A los tweets del archivo JSON se les asignará el valor de cada palabra haciendo coincidir con el diccionario léxico. Como una limitación de las palabras en el diccionario léxico que no puede asignar un valor a cada palabra de los tweets.

Relación con la investigación: La minería de texto será utilizado para categorizar los textos extraídos desde las redes sociales, así mismo aplicar minería de opinión para identificar la reacción de los usuarios ante ciertos eventos que podrían ser: positivo, negativo y neutro.

En la Tabla N° 3 se podrá un observar un resumen de los principales antecedentes encontrados durante el desarrollo de la investigación, así como también su relación.

Tabla N° 3: Cuadro de antecedentes de la investigación

	<b>Antecedente N°1</b>	<b>Antecedente N°2</b>	<b>Antecedente N°3:</b>	<b>Antecedente N°4</b>
Título	Business Intelligence Model to Analyze Social Media Information	Social Business Intelligence: A New Perspective for Decision Makers	Applications of business intelligence and analytics in social media marketing	Twitter Sentiment Analysis
Tipo de Documento	Artículo Procedia Computer Science	Artículo Procedia - Social and Behavioral Sciences	Artículo International Journal of Information Management	Artículo Procedia Computer Science
Objetivo	Analizar contenido de las redes sociales a través de la minería de texto para obtener información relevante para la toma decisiones.	Desarrollar inteligencia social basada en información obtenida de redes sociales a través de la recopilación, clasificación e integración de datos internos y externos	Utilizar las grandes cantidades de datos de redes sociales para captar las percepciones de los consumidores y examinar cómo la Inteligencia de negocios permite a las empresas crear valor y mantener una ventaja competitiva	Realizar minería de opinión sobre textos que implica el estudio computacional de los sentimientos, opiniones y emociones expresadas en párrafos de texto
Metodología	Investigación Descriptiva	Investigación Descriptiva	Investigación Descriptiva	Investigación Descriptiva
Resultados	El algoritmo de prueba de Naive Bayes con Cross Validation con los datos de la tercera prueba en la tercera iteración es la que	Se demuestra que las personas se sienten cómodas intercambiando datos y puntos de vista a	La Inteligencia de Negocios y la Analítica son retos que tienen las empresas que buscan adoptar un enfoque reflexivo y holístico para lograr el aprovechamiento de los	A cada uno de los tweets se les asignará el valor de cada palabra haciendo coincidir

	tiene la mejor precisión	través de entornos sociales.	datos de las redes sociales	con el diccionario léxico.
Relación con la Investigación	Clasificar y categorizar con alta precisión el contenido publicado por usuarios en las redes sociales.	Escuchar a las redes sociales, categorizarlos y transformarlos para obtener conocimiento preciso del comportamiento del consumidor.	La Inteligencia de Negocios Social y la Analítica son herramientas que ayudarán a las empresas en la creación de valor para sus negocios y para mantener una ventaja competitiva sobre su competencia.	La minería de texto será utilizada para categorizar texto extraído de las redes sociales, así como la minería de opinión para identificar reacciones: positivo, negativo y neutro.

Fuente: Elaboración Propia (2020)

## 2.1.1. El sector asegurador en el Perú

### 2.1.1.1. Análisis del sector asegurador

#### 2.1.1.1.1. Empresas de seguros

Según cifras estadísticas de la Superintendencia Nacional de Banca Seguros y AFP a diciembre del 2020, como se puede visualizar en la Figura N° 13, en el Perú operan 18 empresas de Seguros. Las empresas de seguros se pueden agrupar en 3 categorías, de acuerdo con los ramos en los cuales están dedicadas.

- 7 empresas están dedicadas a Ramos Generales y de Vida
- 7 empresas exclusivamente a Ramos Generales
- 4 empresas exclusivas a Ramos de Vida

Figura N° 13: Empresas de Seguros Autorizadas (Al 31 de diciembre del 2020)

Ramos Generales y de Vida	Ramos Generales	Ramos de Vida
Chubb Perú	La Positiva	La Positiva Vida
Cardif	Mapfre Perú	Mapfre Perú Vida
Interseguro	Secrex	Vida Cámara
Rímac	Insur	Ohio Nacional Vida
Pacífico Seguros	Qualitas	
Protecta	Avia Perú	
Crece Seguros	Liberty	

Fuente: Superintendencia Nacional de Banca Seguros y AFP

Por otro lado, en la Figura N° 14 podemos apreciar, según los resultados de la Superintendencia de Banca Seguros y AFP al 31 de diciembre de 2020 se puede visualizar que la participación de mercado del sector asegurador está compuesta de la siguiente manera: 30.64% para Rímac, 26.06% para Pacífico Seguros, 8.15% para La Positiva y 8.14% para Mapfre, las cuales son las empresas más grandes del sector asegurador en el Perú.

Figura N° 14: Ranking de Primas de Seguros Netas por Ramos Al 31 de diciembre del 2020 (En Miles de Soles)

**Total Primas de Seguros**

	Empresas	Monto	Participación (%)	Porcentaje Acumulado
1	Rimac	4 295 949	30.64	30.64
2	Pacifico Seguros	3 653 383	26.06	56.69
3	La Positiva	1 143 158	8.15	64.85
4	Mapfre Perú	1 141 262	8.14	72.99
5	Interseguro	743 550	5.30	78.29
6	La Positiva Vida	711 845	5.08	83.37
7	Mapfre Perú Vida	697 653	4.98	88.34
8	Protecta	408 647	2.91	91.26
9	Cardif	273 425	1.95	93.21
10	Chubb Seguros	236 977	1.69	94.90
11	Ohio National Vida	189 133	1.35	96.25
12	Creceer Seguros	127 516	0.91	97.15
13	AVLA Perú	104 952	0.75	97.90
14	Secrex	92 341	0.66	98.56
15	Vida Cámara	78 542	0.56	99.12
16	Insur	68 462	0.49	99.61
17	Qualitas	34 839	0.25	99.86
18	Liberty Seguros	19 800	0.14	100.00
<b>Total</b>		<b>14 021 434</b>	<b>100.00</b>	

Fuente: Superintendencia Nacional de Banca Seguros y AFP

- **Producción**

Según la Figura N° 16 al 31 de diciembre del 2020, las Primas de Seguros Netas (PSN) anualizadas ascendieron a S/ 14,021,4 millones, lo que representa una caída de 0,7% con respecto a diciembre de 2019.

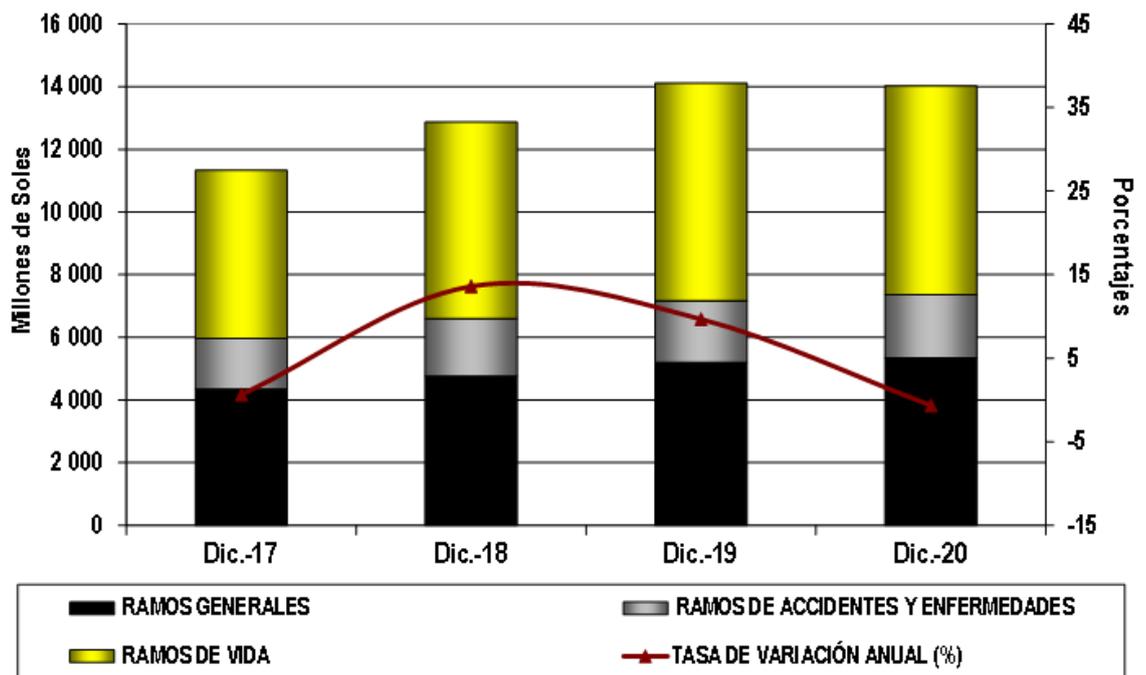
Por otro lado, se observa en la Figura N° 15, la composición de la producción estuvo liderada por los Ramos Generales con una participación de 38.1%, con incremento de 1.3 puntos porcentuales a la registrada en diciembre de 2019. A su vez, los Seguros de Vida registraron una participación de 33.7%, menor en 0,4 p.p. a la observada doce meses antes.

Figura N° 15: Estructuras de las Primas de Seguros Netas, por Ramos Al 31 de diciembre del 2020 (En porcentajes)

	Dic-17	Dic-18	Dic-19	Dic-20
Ramos Generales	38,3	37,0	36,8	38,1
Accidentes y Enfermedades	14,4	14,2	14,0	14,5
Seguros de Vida	30,0	33,5	34,1	33,7
Seguros del SPP	17,2	15,3	15,0	13,8
TOTAL	100,0	100,0	100,0	100,0

Fuente: Superintendencia Nacional de Banca Seguros y AFP

Figura N° 16: Evolución de las Primas de Seguros Netas por Ramos Al 31 de diciembre del 2020



Fuente: Superintendencia Nacional de Banca Seguros y AFP

Como se puede apreciar en la Figura N° 17, las primas netas según ramos se componen de: los Seguros Generales con un 38.09% de la producción en Prima Neta, seguido por los Seguros de Vida con un 33.68%, Seguros de Accidentes y Enfermedades con un 14.46% y finalmente el Sistema Privado de Pensiones con un 13.78%.

Figura N° 17: Primas Según Ramos Al 31 de diciembre del 2020 (En Miles de Soles)

<b>Ramos</b>	<b>Total</b>	<b>% Participación</b>
RAMOS GENERALES	5 340 241	38.09%
RAMOS DE ACCIDENTES Y ENFERMEDADES	2 026 911	14.46%
RAMOS DE VIDA		
Seguros de Vida	4 722 209	33.68%
Seguros del Sistema Privado de Pensiones	1 932 073	13.78%
<b>TOTAL</b>	<b>14 021 434</b>	<b>100.00%</b>

Fuente: Superintendencia Nacional de Banca Seguros y AFP

- **Ramos de No-Vida (Ramos Generales y Seguros de Accidentes y Enfermedades)**

Según la Figura N° 18, las Primas de Seguros Netas (PSN) anualizadas de los Ramos Generales ascendieron a S/ 5,340,1 millones, lo que representó un incremento de 2.8% con respecto a diciembre de 2019. Dentro de este ramo, destacan los incrementos de las PSN anualizadas asociadas a los siguientes riesgos:

- Incendio y Líneas Aliadas (23.3%)
- Caucciones (16.3%)
- Responsabilidad Civil (13.8%)
- Terremoto (5.9%)
- Robo y Asalto (5.2%)

Asimismo, se resaltan las caídas de las PSN anualizadas asociadas a los siguientes riesgos:

- Multiseguros (-32.4%)

- Comprensivo contra Deshonestidad (-10.0%)
- Vehículos (-12.8%)
- Aviación (-7.37%)

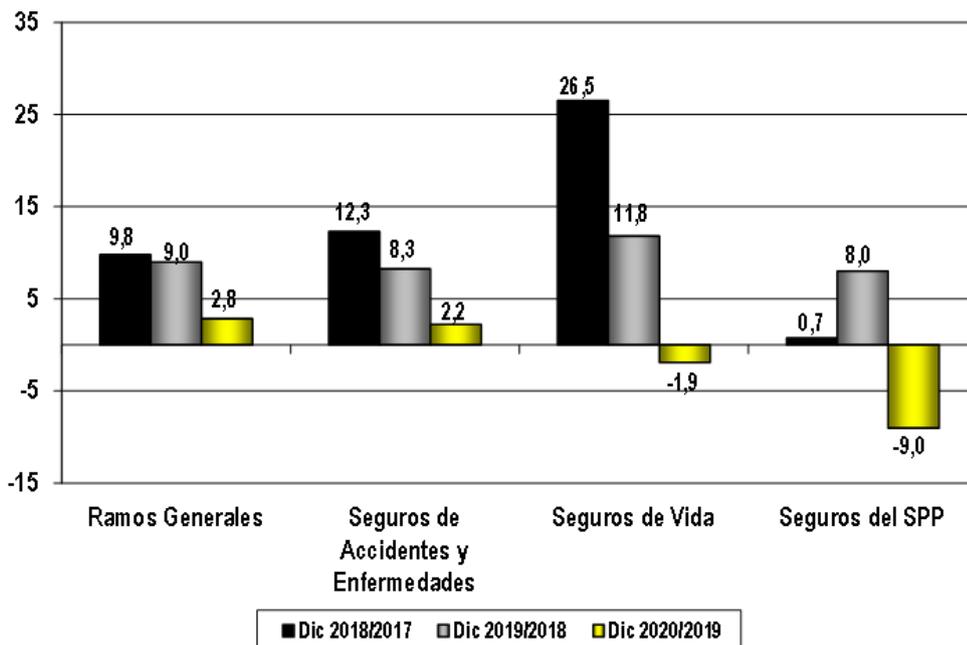
A su vez, las PSN anualizadas de los Seguros de Accidentes y Enfermedades registraron un incremento de 2.2%, alcanzando un volumen de primas de S/ 2,026,9 millones al cierre del año 2020. Dentro de este ramo, destaca el incremento de las PSN anualizadas de los riesgos de:

- Asistencia Médica (8.7%)
- SOAT (1.9%)

Así como se puede observar las caídas en los siguientes riesgos de:

- Accidentes Personales (-19.8%)
- Escolares (-47.5%)

Figura N° 18: Variación anual de Primas de Seguros Netas, por Ramos (En Porcentaje)



Fuente: Superintendencia Nacional de Banca Seguros y AFP

- **Ramos de Vida (Seguros de Vida y del Sistema Privado de Pensiones-SPP)**

Las PSN anualizadas asociadas al Ramo de Vida ascendieron a S/ 6,654,3 millones al 31 de diciembre del 2020, lo que representa una disminución de 4,1% en relación con lo registrado doce meses atrás. Al respecto, las PSN asociadas a los Seguros de Vida presentaron una variación negativa de 1,9%, cayendo hasta un monto de S/ 4,722,2 millones; mientras que las PSN anualizadas de los Seguros del SPP cayeron en 9,0% con relación a diciembre de 2019 hasta un monto de S/ 1,932,1 millones.

En lo referente a los Seguros de Vida, destacan las caídas de las PSN anualizadas asociadas a los riesgos de:

- Vida Individual de Corto Plazo (-31,8%)
- Vida Individual de Largo Plazo (-7,7%)
- Renta Particular (-1,6%)
- SCTR (-1,5%)

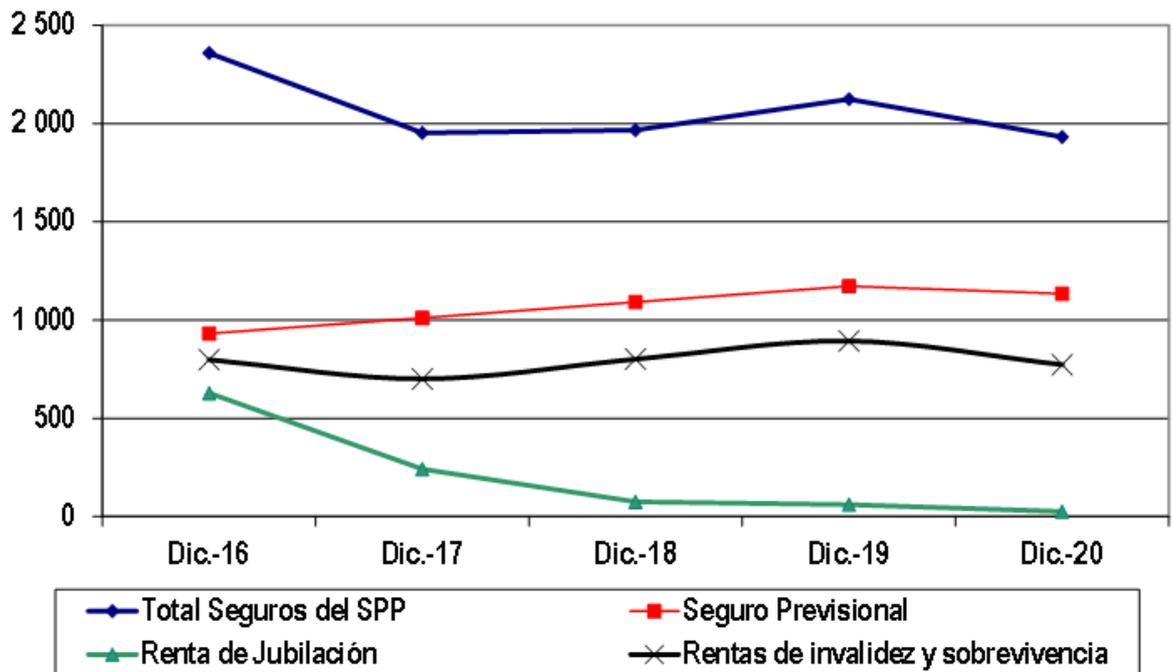
Por otra parte, en el caso del riesgo de Desgravamen se registró un incremento de 1,5%.

Como podemos observar en la Figura N° 19, en relación con los Seguros del SPP, se resaltan las caídas de las PSN anualizadas relacionadas a los riesgos de:

- Rentas de Jubilados (-56,8%)
- Pensiones de Invalidez (-22,3%)

- Pensiones de Sobrevivencia (-4,6%)
- Seguros Previsionales (-3,2%)

Figura N° 19: Primas de Seguros Netas de Seguros del SPP (En millones de Soles)



Fuente: Superintendencia Nacional de Banca Seguros y AFP

#### 2.1.1.1.2. Empresas de corredores de seguros

Los corredores de seguros son empresas intermediarias que comercializan los productos de las compañías aseguradoras y por el cual reciben un porcentaje de comisión establecida por estas últimas.

Las empresas de corredores de seguros trabajan como enlace entre los clientes que necesitan un seguro y las aseguradoras que ofrecen cobertura. Identifican las necesidades de los clientes, y luego contactan con las aseguradoras que tienen más probabilidades de adaptarse a dichas necesidades. La misión del Corredor de Seguros radica en identificar las

necesidades del cliente, evaluar los riesgos y peligros a los que se enfrentan y examinar las políticas existentes de las que gozan. (APECOSE, 2020)

De acuerdo con la base de datos de la Superintendencia de Banca Seguros y AFP que podemos visualizar en la Figura N° 20. Al año 2019 existen 310 empresas corredoras de seguros encabezadas por Marsh Rehder, Contacto, Corredores de Seguros Fallabella y JLT. Además, podemos encontrar 726 corredores de Seguros como Personas Naturales.

Figura N° 20: Información Financiera de Corredores de Seguros Al 31 de diciembre de 2019 (En Soles)

N°	N° REG.	RAZON SOCIAL	INGRESOS OPERATIVOS
1	J0045	MARSH REHDER S.A.C. CORREDORES DE SEGUROS (**)	191 062 910
2	J0359	CONTACTO CORREDORES DE SEGUROS S.A.	51 677 220
3	J0658	CORREDORES DE SEGUROS FALABELLA S.A.C.	43 432 699
4	J0387	JLT PERÚ CORREDORES DE SEGUROS S.A. (**)(***)	37 391 730
5	J0414	CONSEJEROS Y CORREDORES DE SEGUROS S.A.	25 608 511
6	J0569	AON PERU CORREDORES DE SEGUROS S.A. (**)	24 309 670
7	J0377	WILLIS CORREDORES DE SEGUROS S.A. (**)	20 099 000
8	J0288	LA PROTECTORA CORREDORES DE SEGUROS S.A.	19 092 025
9	J0671	SIFUENTES OLAECHEA CORREDORES DE SEGUROS S.A.C.	15 496 511
10	J0394	DETEC S.A. ASESORES Y CORREDORES DE SEGUROS	13 486 468
11	J0833	ARTHUR J. GALLAGHER PERU CORREDORES DE SEGUROS S.A.	12 373 662

Fuente: Superintendencia Nacional de Banca Seguros y AFP

### 2.1.1.2. Perfil del consumidor de seguros

De acuerdo con una encuesta de Datum, realizada del 10 de noviembre de 2011 al 11 de diciembre de 2012 a una muestra de 1,640 personas, el 54% habría pagado por algún seguro propio o para otra persona. Asimismo, el 50% indicó que les resulta difícil comprender las condiciones del seguro y el 66% manifestó que prefiere no pagar un seguro.

Por otra parte, en una muestra de 830 personas encuestadas durante el mismo periodo, el 85% indican que tienen seguros vigentes. En el caso de las Pólizas de Seguro, de una muestra de 400 encuestados pertenecientes al grupo que han pagado algún seguro propio, el 77% manifiesta que han recibido el contrato, siendo un 58% de estos quienes han leído el documento. Con respecto a los Corredores de Seguros, el 37% indica que ha tenido uno.

En el Perú, la necesidad de contratar seguros se vio a fines del siglo XIX. Con el descubrimiento y auge del guano, nuestro Estado emprendió la construcción de ferrocarriles y fue necesario movilizar grandes cantidades de dinero y productos a otros puertos y ciudades. Estos bienes, debían estar adecuadamente protegidos.

Entre las primeras compañías de seguros en el Perú, los historiadores destacan a La Providencia (1863), La Paternal (1867), la Compañía Internacional de Seguros (1895) y la Compañía De Seguros Rímac (1896). Las dos últimas fueron las más reconocidas de fines del siglo XIX y posteriormente, se fusionaron convirtiéndose en lo que hoy conocemos como Rímac Internacional Compañía de Seguros y Reaseguros. Años más tarde ingresan La Compañía De Seguros Italia (1896) y la Sociedad La Positiva Asociación Mutua de Previsión (1937), hoy La Positiva. (Asociación Peruana de Empresas de Seguros, 2018)

## **2.2. MARCO TEÓRICO**

### **2.2.1. Sistemas de información**

Un sistema de información puede definirse técnicamente como un conjunto de componentes interrelacionados que recopilan (o recuperan), procesan, almacenan y distribuyen información para respaldar la toma de decisiones y el control en una organización. Además de apoyar la toma de decisiones, la coordinación y el control, los sistemas de información también pueden ayudar a los gerentes y trabajadores a analizar problemas, visualizar temas complejos y crear nuevos productos. (Laudon & Laudon, 2014)

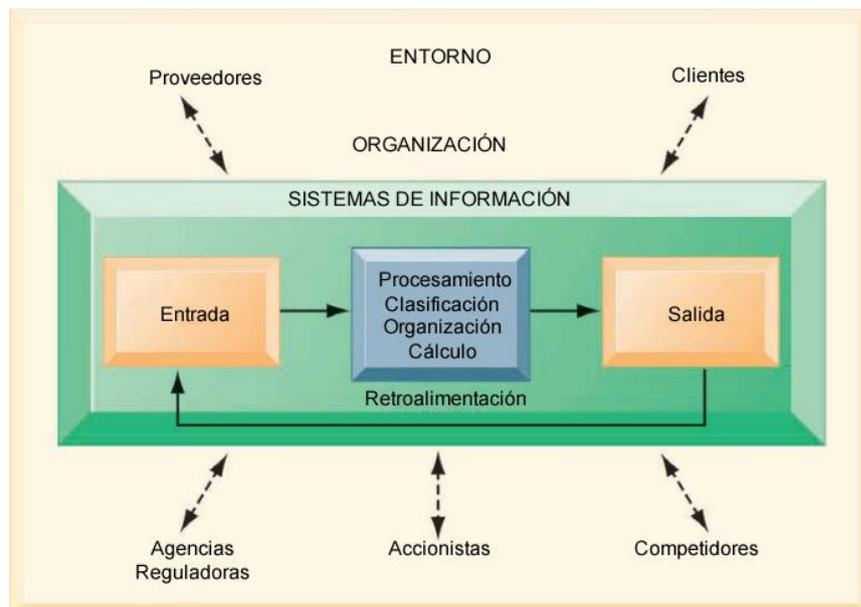
Los sistemas de información contienen información sobre personas, lugares y cosas importantes dentro de la organización o en el entorno que la rodea. Por información nos referimos a datos que han sido moldeados en una forma significativa y útil para los seres humanos. Los datos, por el contrario, son flujos de hechos en bruto que representan eventos que ocurren en las organizaciones o en el entorno físico antes de que se hayan organizado y clasificado en una forma que las personas puedan comprender y utilizar. (Laudon & Laudon, 2014)

En un sistema de información existen tres actividades que producen información para tomar decisiones, controlar operaciones, analizar problemas y crear nuevos productos o servicios. Estas actividades son entrada, procesamiento y salida (Figura N° 21). La entrada captura o recopila datos sin procesar desde dentro de la organización o desde su entorno externo. El procesamiento convierte esta entrada en bruto en una

forma significativa. La salida transfiere la información procesada a las personas que la utilizarán o a las actividades para las que se utilizará. (Laudon & Laudon, 2014)

Los sistemas de información también requieren retroalimentación, que es un resultado que se devuelve a los miembros apropiados de la organización para ayudarlos a evaluar o corregir la etapa de entrada.

Figura N° 21: Funciones de un Sistema de Información



Fuente: Management Information Systems Managing, 13<sup>th</sup> Edition

### 2.2.1.1. Dimensiones de los sistemas de información

Para comprender completamente los sistemas de información, debe comprender las dimensiones más amplias de: organización, administración y tecnología de la información de los sistemas (Figura N° 22) y su poder para brindar soluciones a los desafíos y problemas en el entorno empresarial. Nos referimos a esta comprensión más amplia de los sistemas de información, que abarca la comprensión de las dimensiones organizativas y de gestión de

los sistemas, así como las dimensiones técnicas de los sistemas, como alfabetización en sistemas de información. La alfabetización informática, por el contrario, se centra en el conocimiento de la tecnología de la información. (Laudon & Laudon, 2014)

Figura N° 22: Dimensiones de un Sistema de Información



Fuente: Management Information Systems Managing, 13<sup>th</sup> Edition

#### **2.2.1.2. Tipos de sistemas de información**

La necesidad de información de una organización es variados y diverso. Esto es debido a que existen distintas especialidades y niveles jerárquicos con intereses y responsabilidades muy diferentes.

Una organización tiene procesos de soporte de sistemas para cada una de las funciones comerciales principales: ventas y marketing, manufactura y producción, finanzas y contabilidad, y recursos humanos. Los sistemas funcionales que operan de forma independiente entre sí se están

convirtiéndose en una cosa del pasado porque no pueden compartir información fácilmente para soportar los procesos de negocio multifuncionales. Muchos han sido reemplazados por sistemas multifuncionales a gran escala que integran las actividades de los procesos comerciales y las unidades organizativas relacionados. (Laudon & Laudon, 2014)

Una empresa comercial tiene sistemas para apoyar a diferentes grupos o niveles de gestión. Estos sistemas incluyen sistemas de procesamiento de transacciones y sistemas para inteligencia de negocios.

Una empresa tiene sistemas dirigidos a los diferentes grupos o niveles de gestión. Estos sistemas incluyen sistemas de: procesamiento de transacciones y sistemas para inteligencia de negocios.

#### **2.2.1.2.1. Sistemas de procesamiento de transacciones**

El objetivo de los sistemas de procesamiento de transacciones es capturar y procesar datos sobre las transacciones de negocios que se realizan, diariamente,

en la empresa. Las transacciones son hechos o actividades que se llevan a cabo en la empresa, y que le aportan nueva información.

Los sistemas de procesamiento de transacciones (TPS) proporcionan este tipo de información. Un sistema de procesamiento de transacciones es un sistema computarizado que realiza y registra las transacciones de rutina diaria necesarias para realizar negocios, como el ingreso de órdenes de

venta, reservas de hotel, nómina, mantenimiento de registros de empleados y envío. (Laudon & Laudon, 2014)

#### **2.2.1.2.2. Sistemas de Inteligencia de Negocios**

Las empresas tienen sistemas de inteligencia de negocios para brindar información para la toma de decisiones. La inteligencia de negocios es un término contemporáneo para los datos y las herramientas de software para organizar, analizar y proporcionar acceso a los datos que ayuda a los gerentes y otros usuarios de negocios a tomar decisiones más informadas. (Laudon & Laudon, 2014)

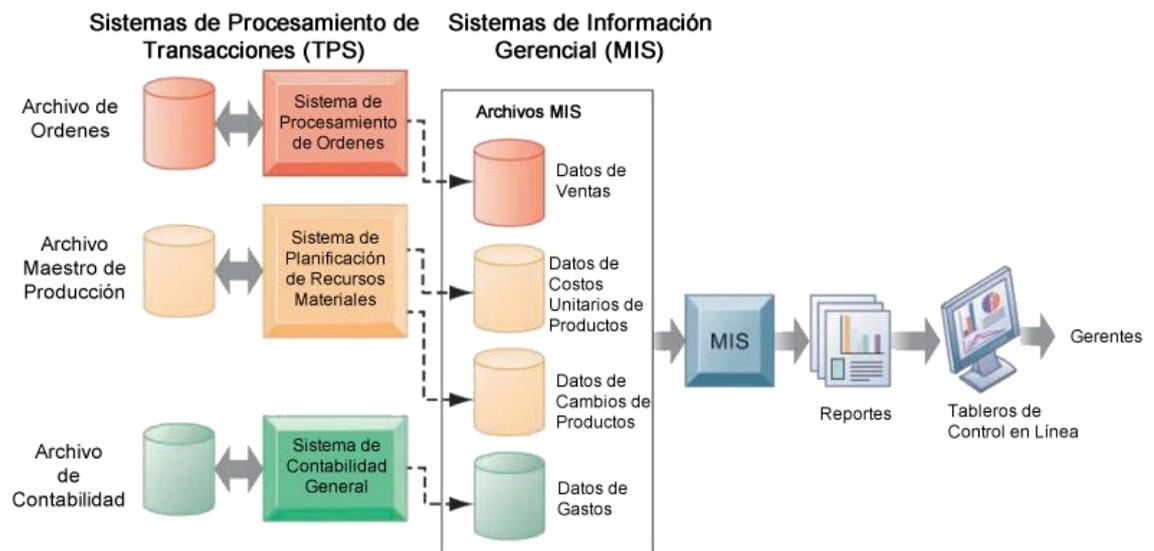
- **Sistemas de Información Gerencial (MIS)**

El término Sistemas de Información Gerencial (MIS) también designa una categoría específica de sistemas de información al servicio de la gerencia media. MIS proporciona a los mandos intermedios informes sobre el desempeño actual de la organización. Esta información se utiliza para monitorear y controlar el negocio y predecir el desempeño futuro. (Laudon & Laudon, 2014)

MIS resume e informa sobre las operaciones de negocio de la empresa utilizando datos proporcionados por los sistemas de procesamiento de transacciones (TPS) los cuales se comprimen y generalmente se presentan en informes que se producen en un horario regular. Estos reportes se entregan en línea. La Figura N° 23 muestra cómo un MIS transforma los datos a nivel de transacción: inventario, producción y contabilidad en

archivos MIS que luego se utilizan para proporcionar reportes a los gerentes.  
(Laudon & Laudon, 2014)

Figura N° 23: Cómo los Sistemas de Información Obtienen datos del TPS



Fuente: Management Information Systems Managing, 13<sup>th</sup> Edition

Los MIS proporcionan respuestas a preguntas que se han especificado de antemano y tienen un procedimiento predefinido para responderlas. Estos sistemas generalmente no son flexibles y tienen poca capacidad analítica. La mayoría de los MIS utilizan rutinas simples, como resúmenes y comparaciones, en contraposición a modelos matemáticos sofisticados o técnicas estadísticas. (Laudon & Laudon, 2014)

- **Sistemas de Apoyo a la Toma de Decisiones (DSS)**

Los sistemas de apoyo a la toma de decisiones (DSS) se enfocan en problemas que son únicos y que cambian rápidamente, para los cuales el procedimiento para llegar a una solución puede no estar completamente

predefinido de antemano. Intentan responder preguntas como estas: ¿Cuál sería el impacto en los programas de producción si duplicamos las ventas en el mes de diciembre? ¿Qué pasaría con nuestro retorno de la inversión si el programa de una fábrica se retrasara seis meses? (Laudon & Laudon, 2014)

Aunque los DSS utilizan información interna de TPS y MIS, a menudo aportan información de fuentes externas, como los precios actuales de las acciones o los precios de los productos de la competencia. Estos sistemas son empleados por gerentes y analistas de negocios que desean utilizar análisis y modelos sofisticados para el análisis de datos. (Laudon & Laudon, 2014)

- **Sistemas de Apoyo a Ejecutivos (ESS)**

Son sistemas de información a nivel estratégico y ayudan a la alta dirección a la toma de decisiones. Abordan decisiones no rutinarias que requieren juicio, evaluación y conocimiento porque no existe un procedimiento acordado para llegar a una solución. ESS presenta gráficos y datos de muchas fuentes a través de una interfaz que es fácil de usar para los altos directivos. A menudo, la información se entrega a los altos ejecutivos a través de un portal, que utiliza una interfaz web para presentar contenido empresarial personalizado e integrado. (Laudon & Laudon, 2014)

### **2.2.2. Sistemas de Inteligencia de Negocios (BI)**

Uno de los activos más importantes de cualquier organización es su información. Este activo se usa casi siempre para dos propósitos:

mantenimiento de los sistemas de soporte a las operaciones de negocio y para la toma de decisiones. (Kimball & Ross, 2013)

Los usuarios de un sistema de operaciones de negocio son los que realizan las tareas y actividades diarias como son las de: registrar pedidos, registrar nuevos clientes, consultar el estado de una orden y registrar reclamos de los clientes. Los sistemas de operaciones de negocio han sido optimizados para procesar las transacciones rápidamente. (Kimball & Ross, 2013)

Por otro lado, los usuarios de un sistema Inteligencia de Negocios (BI), son los que, observan a la organización para poder evaluar su rendimiento. Los sistemas de Inteligencia de Negocios (BI) son sistemas que han sido optimizados para responder a consultas de alto rendimiento, ya que las consultas de los usuarios requieren de cientos de miles de transacciones se busquen, agrupen y acumulen un conjunto de datos. (Kimball & Ross, 2013)

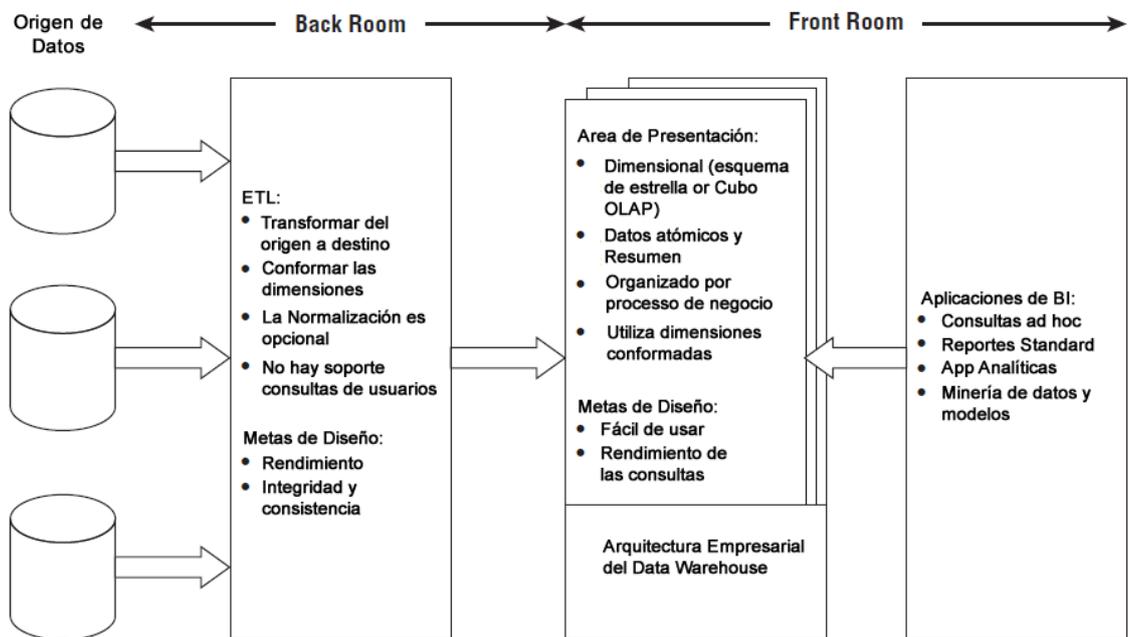
Los sistemas de Inteligencia de Negocios antes de considerar la tecnología y las herramientas que se utilizará, debe considerar las necesidades de la empresa. Con las necesidades del negocio ya definidas, luego podemos trabajar hacia atrás a través de los diseños lógicos y luego físicos, junto con las decisiones sobre tecnología y herramientas. (Kimball & Ross, 2013)

#### **2.2.2.1. Arquitectura Inteligencia de Negocios (BI)**

Como se puede visualizar en la Figura N° 24, dentro de todo proyecto de Inteligencia de Negocios (BI) existen cuatro componentes separados y

distintos: Sistemas de Soporte a las Operaciones de Negocio, ETL (Extracción Transformación y Carga), Área de Presentación y Aplicaciones de Inteligencia de Negocios (BI), los cuales son las aplicaciones encargadas de realizar la explotación de los datos.

Figura N° 24: Arquitectura Inteligencia de Negocios (BI)



Fuente: The Data Warehouse Toolkit, 3<sup>rd</sup> Edition

### 2.2.2.1.1. Sistema de soporte operaciones de negocio

Son los sistemas que dan soporte y capturan el día a día de las operaciones del negocio. Las prioridades de estos sistemas son la lograr una alta performance y disponibilidad. Las consultas contra los sistemas de operaciones de negocio son limitadas, se consulta un registro a la vez y mantienen pocos datos históricos

#### **2.2.2.1.2. Extracción, transformación y carga (ETL)**

El área de almacenamiento del almacén de datos (Data Warehouse), es tanto un área de almacenamiento como un conjunto de procesos a los que se hace referencia comúnmente como: Extracción, Transformación y Carga (ETL).

El área de almacenamiento de datos se encuentra entre los sistemas de operaciones de negocio y el área de presentación de datos. Es algo similar a la cocina de un restaurante, donde los productos crudos se transforman en una buena comida. En el almacén de datos, los datos de las operaciones de negocio sin procesar se transforman para la consulta y el consumo de los usuarios.

#### **2.2.2.1.3. Área de presentación**

El área de presentación de datos es donde se organizan, almacenan y ponen a disposición los datos para que los usuarios, los generadores de reportes y otras aplicaciones analíticas y de explotación realicen consultas directas.

#### **2.2.2.1.4. Aplicaciones de Inteligencia de Negocios (BI)**

El componente principal de la arquitectura de BI es la Aplicación de Inteligencia de Negocios (BI). El término Aplicación de BI se refiere libremente a la gama de capacidades proporcionadas a los usuarios de negocios para aprovechar el área de presentación para la toma de decisiones analíticas.

Por definición, todas las aplicaciones de Inteligencia de Negocios consultan los datos en el área de presentación de BI. La consulta tiene como objetivo de usar los datos para mejorar la toma de decisiones.

Una aplicación de Inteligencia de Negocios puede ser tan simple como una herramienta de consulta ad hoc o tan compleja como una sofisticada aplicación de minería de datos. Las herramientas de consulta ad hoc, son tan poderosas y solo pueden ser entendidas y utilizadas de manera efectiva por un pequeño porcentaje de usuarios de negocios

Lo más probable es que los usuarios de negocios accedan a los datos a través de aplicaciones y plantillas preconstruidas basadas en parámetros que no requieren que los usuarios construyan las consultas directamente.

Algunas de las aplicaciones de Inteligencia de Negocios son más sofisticadas, como las herramientas de pronósticos, que permitir subir los resultados de las consultas a los Sistemas de Soporte a las Operaciones de Negocio, al ETL o al área de presentación. (Kimball & Ross, 2013)

#### **2.2.2.2. Modelamiento dimensional**

El modelamiento dimensional está ampliamente aceptado como la técnica preferida para presentar datos analíticos porque aborda dos requisitos simultáneos: datos que sean fácilmente entendidos por los usuarios de negocio y una alta performance en consultas rápidas. (Kimball & Ross, 2013)

Durante más de cinco décadas, las organizaciones de TI, los consultores de TI y los usuarios de negocios se han inclinado hacia una estructura

dimensional simple para satisfacer la necesidad humana de la simplicidad. La simplicidad es fundamental porque garantiza que los usuarios puedan comprender fácilmente los datos, así como también permite que el software navegue y entregue resultados de manera rápida y eficiente. (Kimball & Ross, 2013)

Aunque los modelos dimensionales a menudo se instancian sobre sistemas RDBMS (Relational Database Management System), son bastante diferentes de los modelos de tercera forma normal (3NF) que buscan eliminar la redundancia de datos. Las estructuras 3NF normalizadas dividen los datos en muchas entidades, cada una de las cuales se convierte en una tabla relacional. (Kimball & Ross, 2013)

La industria a veces se refiere a los modelos 3NF como modelos de entidad-relación (ER). Los diagramas de relación de entidad (diagramas ER o ERD) son dibujos que comunican las relaciones entre tablas. Tanto los modelos 3NF como los modelos dimensionales se pueden representar en diagramas ER (ERD) porque ambos consisten en tablas unidas a través de relacionales. La diferencia entre 3NF y los modelos dimensionales es el grado de normalización. (Kimball & Ross, 2013)

Las estructuras 3NF normalizadas son muy eficientes con las bases de datos OLTP (Online Transactional Processing) porque una transacción de inserción, actualización o eliminación toca la base de datos en un solo lugar. Sin embargo, los modelos normalizados son complicados para ejecutar

consultas de sistemas de Inteligencia de Negocios. Los usuarios no pueden entender, navegar ni recordar modelos normalizados que se asemejen a un mapa del sistema de autopistas, asimismo, la mayoría de los sistemas de gestión de bases de datos relacionales no pueden consultar eficientemente un modelo normalizado, la complejidad de las consultas impredecibles de los usuarios sobre carga a los optimizadores de la base de datos, lo que resulta en un rendimiento pobre de las consultas. (Kimball & Ross, 2013)

#### **2.2.2.2.1. Tabla de hechos**

La tabla de hechos en un modelo dimensional almacena las mediciones de rendimiento resultante de los eventos de los procesos de negocio. Se deben almacenar los datos de medición de bajo nivel resultante de un proceso de negocio en un solo modelo dimensional.

Debido a que los datos de medición son enormes, no deben replicarse en múltiples lugares para múltiples funciones en la empresa. Permitir que los usuarios de negocio accedan a un único repositorio de datos, garantiza el uso de datos consistentes en toda la empresa.

Cada fila en una tabla de hechos corresponde a un evento de medición. Los datos de cada fila se encuentran en un nivel de detalle denominado granularidad, como una fila por producto vendido en una transacción de venta. Uno de los principios básicos del modelado dimensional es que todas las filas de medición en una tabla de hechos deben estar en el mismo nivel de granularidad.

La aditividad en una tabla de hechos es muy importante porque las aplicaciones de BI consultan cientos, miles o incluso millones de filas a la vez, y lo más útil que se puede hacer con tantas filas es sumarlas.

Algunos hechos son a veces semi-aditivos o incluso no aditivos. Los hechos semi-aditivos, como los saldos de cuentas, no se pueden sumar a través de la dimensión tiempo. Los hechos no aditivos, como los precios unitarios, nunca se pueden acumular. Se tienen que contar o promediar o solo se podrán imprimir.

#### **2.2.2.2.2. Tabla de dimensiones**

Las tablas de dimensiones están íntegramente relacionadas a una tabla de hechos. Las tablas de dimensiones contienen el contexto descriptivo asociado con un evento de medición de un proceso de negocio. Describen el "quién, qué, dónde, cuándo, cómo y por qué" asociado con el evento.

Las tablas de dimensiones tienen muchas columnas o atributos, aunque, algunas tablas de dimensiones naturalmente tienen solo cuantos atributos. Las tablas de dimensiones tienen menos filas que las tablas de hechos, pero pueden tener muchas columnas de texto de longitud amplia. Cada dimensión se define mediante una sola clave primaria que sirve como base para la integridad referencial con cualquier tabla de hechos a la que se relaciona.

Los atributos de dimensión sirven como criterios de filtros de consultas, agrupaciones y etiquetas de reportes. En una consulta o solicitud de informe, los atributos se identifican como por palabras. Por ejemplo, cuando un

usuario desea ver las ventas en dólares por marca, la marca debe estar disponible como un atributo de dimensión.

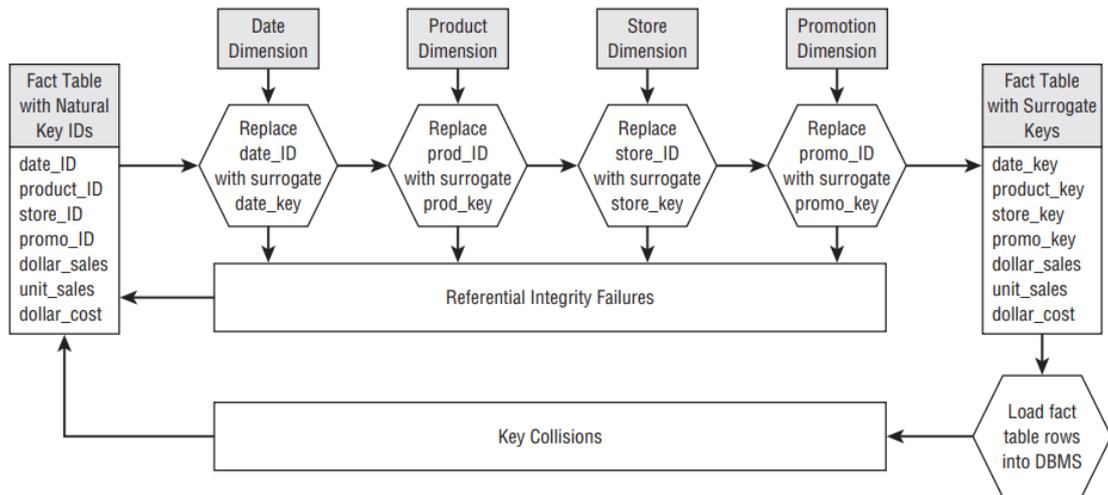
#### **2.2.2.2.3. Claves sustitutas de dimensión**

Toda tabla de dimensión es creada con una columna que sirve como clave principal, esta clave principal no puede ser la clave natural del sistema de soporte a las operaciones de negocios porque habría varias filas de dimensiones relacionadas a esa clave natural cuando se realice un seguimiento a los cambios a lo largo del tiempo, además las claves naturales pueden ser creadas por un sistema fuente que puede ser mal administrada. (Kimball & Ross, 2013)

Ante esta realidad es que se hace necesario que el sistema de inteligencia de negocios (BI) tome el control en la creación de claves primarias de todas las dimensiones del modelo. Como se puede visualizar en la Figura N° 25 en lugar de utilizar las claves naturales o claves de negocio se debe crear claves enteras anónimas conocidas como claves sustitutas para cada dimensión.

Las claves sustitutas de una dimensión son números enteros simples que son generados en secuencia por el motor de base de datos y comienzan generalmente por el número 1. Las claves sustitutas son también conocidas como: clave sintética, identificador de entidad, clave generada por el sistema, número de secuencia de la base de datos, clave técnica o identificador único arbitrario.

Figura N° 25: Reemplazando claves naturales o de negocio por claves sustitutas

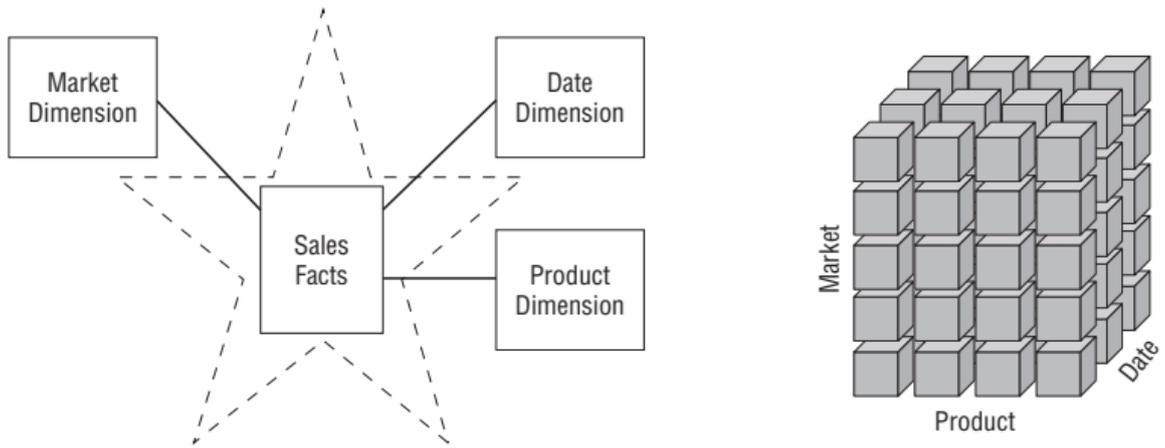


Fuente: The Data Warehouse Toolkit, 3<sup>rd</sup> Edition

#### 2.2.2.2.4. Esquema estrella

Como se puede visualizar en la Figura N° 26, los modelos dimensionales que son desplegados en los sistemas RDBMS se denominan esquemas estrella debido a su parecido con una estructura en forma de estrella. Los modelos dimensionales implementados en entornos de bases de datos multidimensionales se denominan cubos OLAP (Online Analytical Processing). (Kimball & Ross, 2013)

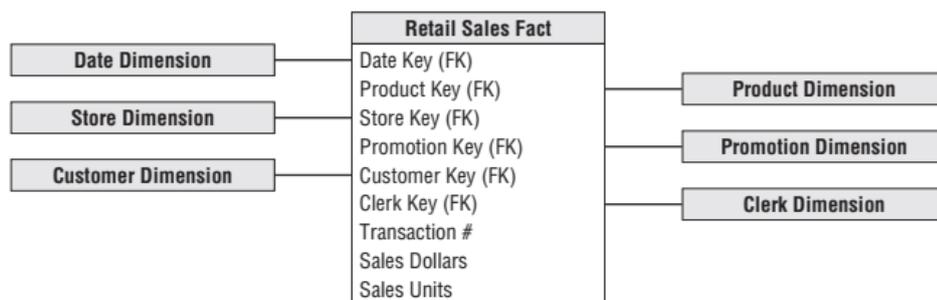
Figura N° 26: Esquema Estrella vs cubos OLAP



Fuente: The Data Warehouse Toolkit, 3<sup>rd</sup> Edition

Cada proceso de negocios está representado por un modelo dimensional que consta de una tabla de hechos que contiene las medidas numéricas del evento rodeadas por un halo de tablas de dimensiones que contienen el contexto textual que era verdadero en el momento en que ocurrió el evento. Esta estructura característica en forma de estrella a menudo se denomina unión en estrella, un término que se remonta a los primeros días de las bases de datos relacionales. (Kimball & Ross, 2013)

Figura N° 27: Tabla de Hechos y Dimensiones relacionadas en un Esquema Estrella



Fuente: The Data Warehouse Toolkit, 3<sup>rd</sup> Edition

Lo primero que hay que notar sobre el esquema estrella es su simplicidad y simetría. Obviamente, los usuarios de negocio se benefician de la simplicidad porque los datos son más fáciles de entender y navegar. El atractivo del diseño de la Figura N° 27 es que es fácilmente reconocido por los usuarios de negocio. (Kimball & Ross, 2013)

La simplicidad de un modelo dimensional también tiene beneficios de rendimiento. Los optimizadores de bases de datos procesan estos esquemas simples con menos combinaciones de manera más eficiente. Un motor de base de datos puede hacer suposiciones sólidas acerca de restringir primero las tablas de dimensiones altamente indexadas y luego atacar la tabla de hechos de una vez con el producto cartesiano de las claves de la tabla de dimensiones que satisfacen las restricciones del usuario. Sorprendentemente, con este enfoque, el optimizador puede evaluar uniones arbitrarias de  $n$  vías a una tabla de hechos en una sola pasada a través del índice de la tabla de hechos. (Kimball & Ross, 2013)

#### **2.2.2.2.5. Diseño dimensional**

De acuerdo con (Kimball & Ross, 2013), las cuatro decisiones clave que se toman durante el diseño de un modelo dimensional incluyen:

- Seleccionar un proceso de negocio
- Declare la granularidad de los datos
- Identificar las Dimensiones
- Identificar los Hechos

Las respuestas a estas preguntas se determinan considerando las necesidades de la empresa junto con la disponibilidad de los datos de origen subyacentes durante las sesiones de modelado colaborativo. Siguiendo el proceso de negocio, la granularidad, la dimensión y las declaraciones de hechos, el equipo de diseño determina los nombres de las tablas y columnas, los valores de dominio de muestra y las reglas empresariales. (Kimball & Ross, 2013)

### **2.2.3. Inteligencia de Negocios Social (SBI)**

De acuerdo con (Muntean, Cabău, & Rînciog, 2014), en los últimos años, las personas se sienten cómodas de intercambiar datos y puntos de vista a través de las redes sociales.

Asimismo, según (Harrysson, Metayer, & Sarrazin, 2019), las empresas pueden desarrollar una "inteligencia social" basada en la información e ideas difundidas a través de las redes sociales por sus empleados, clientes y grupos de interés.

El contenido recopilado de las redes sociales se analizará y procesará para obtener conocimiento valioso, que enriquecerá la visión de la empresa lo cual implícitamente ayudará a tomar mejores decisiones.

El proceso de recopilar datos de las redes sociales, analizarlos para tomar mejores decisiones se conoce como: Inteligencia de Negocios Social (SBI)

(Muntean, Cabău, & Rînciog, 2014), es un nuevo enfoque de Inteligencia de Negocios (BI) enriquecido con "inteligencia social"

### **2.2.3.1. Arquitectura Inteligencia de Negocios Social**

La arquitectura de un proyecto de Inteligencia de Negocios Social (SBI) está compuesto de los componentes de un proyecto de Inteligencia de Negocios tradicional a los que se le ha añadido nuevos componentes como, la conexión a servicios externos a través del consumo de una API, programas rastreadores o robots de extracción de datos textuales, programas clasificadores de texto y programas de descubrimiento de sentimientos en grandes volúmenes de texto.

Como se puede observar en la Figura N° 28, los componentes que se han adicionado a la arquitectura se encuentran en la capa de extracción y son: La API Graph, el rastreador, el programa de Clasificador de Texto y el programa de Análisis de Sentimiento.

- **La API Graph**

Las Interfaz de Programación de Aplicaciones (API) son servicios expuestos por cada una de las plataformas de las redes sociales y pueden ser consumidos por las aplicaciones construidas por empresas o desarrolladores de manera libre y luego podrán integradas a aplicaciones empresariales o aplicaciones de Inteligencia de Negocios Social.

La fuente de datos del proyecto de Inteligencia de Negocios Social (SBI) serán obtenida directamente desde la plataforma de redes sociales:

Facebook. Los datos textuales en bruto serán obtenidos por los rastreadores y luego serán almacenados en las bases de datos para su procesamiento, clasificación y descubrimiento de comportamiento.

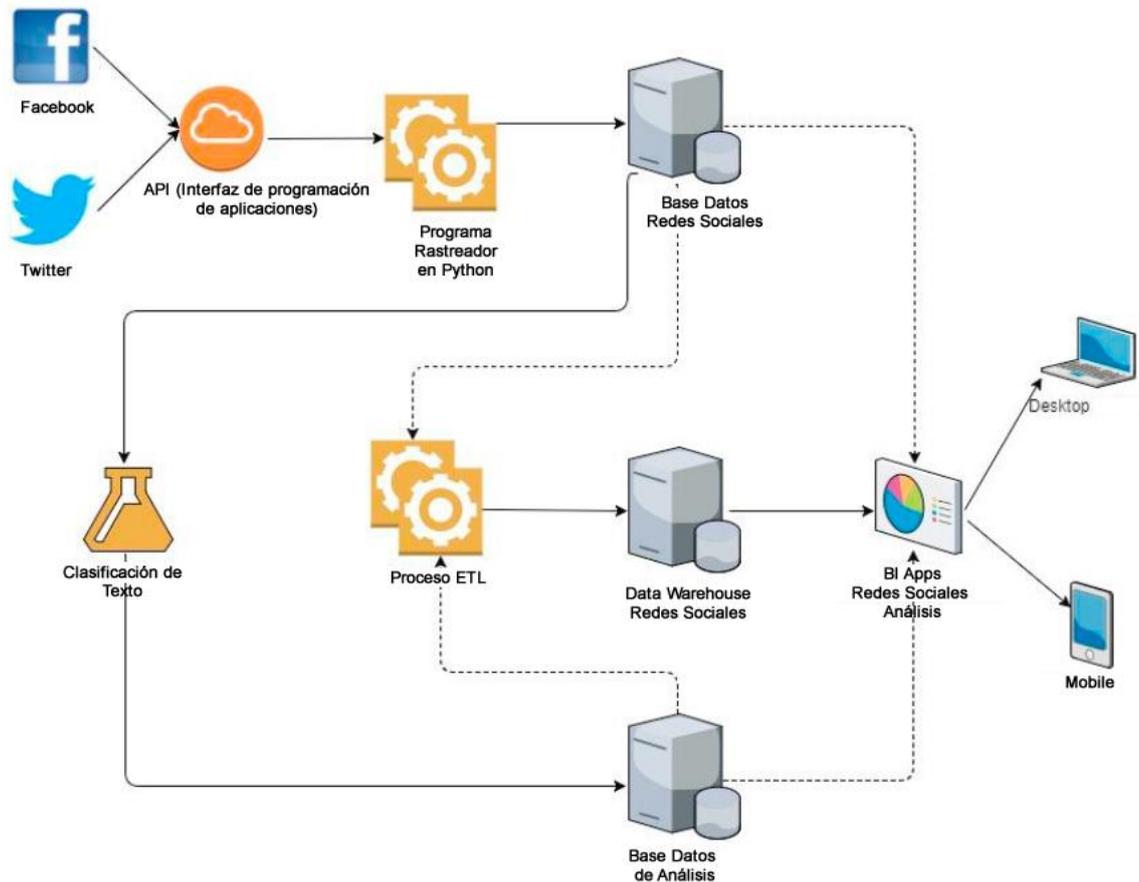
- **Programa Rastreador**

En la presente investigación desarrollaremos programas rastreadores capaces de consumir los servicios expuestos por la API Graph de Facebook utilizando el lenguaje de programación Python, estos programas son también conocidos como robots.

- **Clasificador de Texto**

Para lograr la clasificación y categorización de los textos se le utilizarán algoritmos de minería de texto, el cual es un proceso de colocar un documento o porción de texto en una categoría o clase de acuerdo con las características del documento. Este proceso se realizará través de la Extracción, Transformación y Carga (ETL).

Figura N° 28: Arquitectura Inteligencia de Negocios Social (SBI)



Fuente: Business Intelligence Model to Analyze Social Media Information (Kurnia & Suharjo, 2018)

#### 2.2.4. Machine Learning

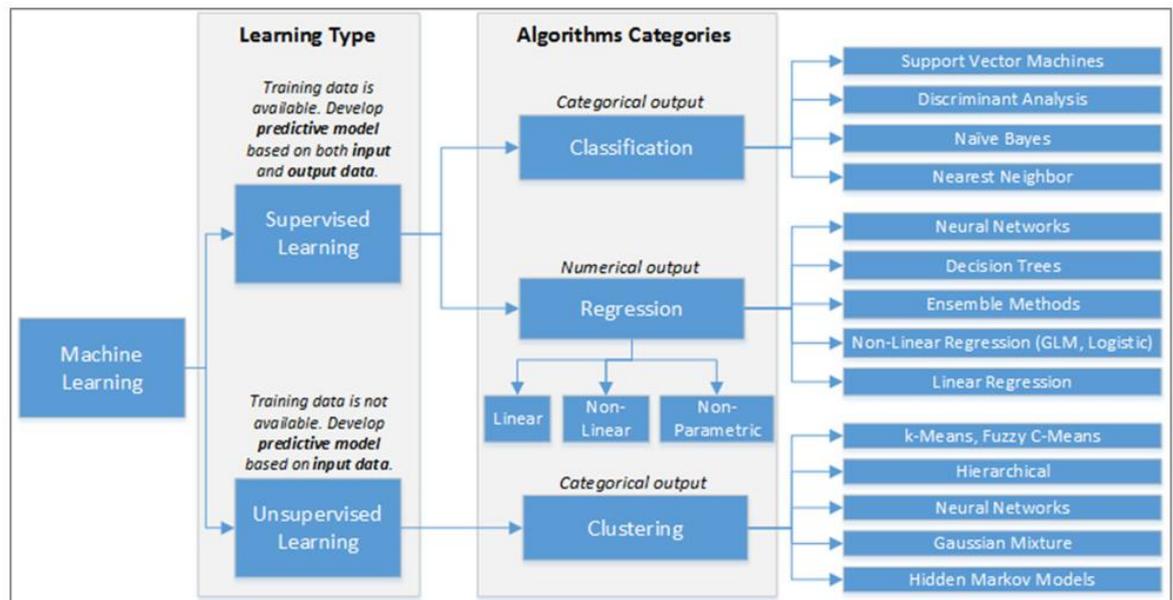
El término “Machine Learning” se define como un conjunto de métodos que pueden detectar automáticamente patrones en los datos y luego usar los patrones descubiertos para predecir datos futuros o para realizar otros tipos de toma de decisiones en condiciones de incertidumbre como son la planificación o la recopilación de más datos. (Murphy, 2012)

Machine Learning tiene como objetivo resolver el problema de tener grandes cantidades de datos con muchas variables, se usa mayormente en áreas

como el reconocimiento de patrones de voz e imágenes; algoritmos financieros como: calificación crediticia, comercio algorítmico; pronóstico de energía: carga, precio y biología: detección de tumores, descubrimiento de fármacos. (Batrinca & Treleaven, 2015)

En la Figura N° 29 podemos visualizar los dos tipos de aprendizaje de Machine Learning y sus categorías de algoritmos.

Figura N° 29: Machine Learning Overview



Fuente: Social media analytics: a survey of techniques, tools and platforms (Batrinca & Treleaven, 2015)

El primer paso para ejecutar los algoritmos de Machine Learning es preparar ejemplos de entrenamiento o datos de muestra. La capacidad de clasificación, regresión o clustering se da en varias formas dependiendo de los tipos de algoritmos de aprendizaje automático y se construye mediante ejemplos preparados, también conocido como el proceso de aprendizaje. Los ejemplos iniciales que se denominan ejemplos de prueba son

clasificados en la tarea de clasificación y los valores de salida son estimados en la tarea de regresión, esto se llama generalización. (Jo, 2019)

Los algoritmos de Machine Learning se aplican a tareas como: la clasificación, la regresión y el clustering. La clasificación se refiere a la asignación de una o algunas de las categorías predefinidas a cada objeto. La regresión es el proceso de estimar un valor de salida o valores de salida que están relacionados de forma lineal o no lineal con los valores de entrada. El clustering significa el proceso de dividir un grupo de objetos en subgrupos de otros similares. (Jo, 2019)

#### **2.2.4.1. Tipos de Machine Learning**

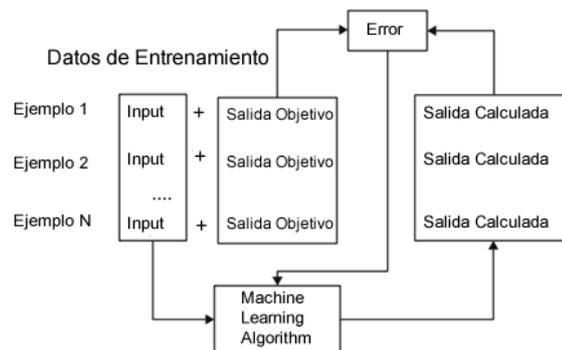
Existen dos tipos de algoritmos de Machine Learning, el aprendizaje supervisado y el aprendizaje no supervisado. El primero se utiliza para la clasificación y la regresión, y el segundo se utiliza para el clustering.

##### **2.2.4.1.1. Aprendizaje supervisado**

El algoritmo de aprendizaje supervisado como se puede observar en la Figura N° 30, asume que todos los datos de entrenamiento están etiquetados con sus propios valores de salida, que se denominan salidas objetivo. Los parámetros del algoritmo de Machine Learning se inicializan de forma aleatoria y los valores de salida de los ejemplos de entrenamiento son calculados con los parámetros a los cuales también se les denomina salidas calculadas. (Jo, 2019)

A la diferencia entre los tipos de salida: las salidas objetivo y las salidas calculadas, se les denomina error y se convierte en la base para corregir los parámetros. El aprendizaje supervisado se refiere al proceso de optimización de los parámetros de Machine Learning para minimizar el error. (Jo, 2019)

Figura N° 30: Aprendizaje Supervisado



Fuente: Text Mining Concepts, Implementation, and Big Data Challenge (Jo, 2019)

En el proceso de aprendizaje, se deben minimizar las diferencias entre las etiquetas de destino y las calculadas. El aprendizaje supervisado se aplica a la clasificación donde las salidas se dan como valores discretos y la regresión donde se dan como valores continuos. K Nearest Neighbor, Naive Bayes, Perceptron y Support Vector Machine son algoritmos típicos de aprendizaje supervisado, pero se pueden modificar para convertirlos en algoritmos no supervisados. (Jo, 2019)

#### 2.2.4.1.1.1. Support Vector Machine (SVM)

El algoritmo Support Vector Machine (SVM) es un grupo de algoritmos de aprendizaje que se utilizan principalmente para tareas de clasificación en datos complicados, como la clasificación de imágenes y el análisis de la estructura de proteínas. (Mertsalov & McCreary, 2019)

El algoritmo Support Vector Machine (SVM) es uno de los algoritmos de Machine Learning supervisado el cual es empleado para resolver varios problemas de clasificación. Tiene sus aplicaciones en análisis de riesgo crediticio, diagnóstico médico, categorización de texto y extracción de información. (Thangaraj & Sivakami, 2018)

Adicionalmente, el algoritmo Support Vector Machine (SVM) se utiliza en innumerables campos de la ciencia y la industria, incluida la biotecnología, la medicina, la química y la informática. Como también ha resultado ser ideal para la categorización de grandes repositorios de texto, como los que se encuentran en prácticamente todas las grandes organizaciones modernas. (Mertsalov & McCreary, 2019)

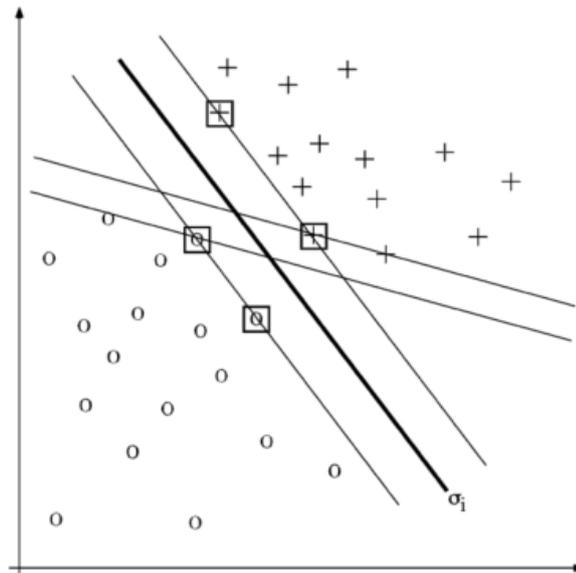
De acuerdo a (Sebastiani, 2002), el algoritmo de aprendizaje supervisado Support Vector Machine (SVM) fue introducido para la Clasificación de Texto (TC) por (Joachims, 1999) y posteriormente fue utilizado por (Drucker, Vapnik, & Wu, 1999), (Dumais, Platt, Heckerman, & Sahami, 1998), (Dumais & Chen, Hierarchical classification of Web content, 2000), (Klinkenberg & Joachims, 2000), (Taira & Haruno, 1999) y (Yang & Liu, 1999).

Son modelos de aprendizaje supervisado que están relacionados con algoritmos de aprendizaje que analizan datos y reconocen patrones, que se utilizan para el análisis de clasificación y regresión. (Batrinca & Treleaven, 2015)

En términos geométricos, puede verse como el intento de encontrar, entre todas las superficies  $\sigma_1, \sigma_2, \dots$  en el espacio  $|T|$ -dimensional que separan los ejemplos de entrenamiento positivos de los negativos (superficies de decisión), el  $\sigma_i$  que separa los positivos de los negativos por el margen más amplio posible, es decir, de manera que la propiedad de separación sea invariante con respecto a la traducción más amplia posible de  $\sigma_i$ . (Sebastiani, 2002)

Esta idea se comprende mejor en el caso en el que los positivos y los negativos son linealmente separables, en cuyo caso las superficies de decisión son  $(|T|-1)$ -hiperplanos. En el caso bidimensional de la Figura N° 31, se pueden elegir varias líneas como superficies de decisión. El método SVM elige el elemento del medio del conjunto "más ancho" de líneas paralelas, es decir, del conjunto en el que la distancia máxima entre dos elementos del conjunto es la más alta. Es de destacar que esta "mejor" superficie de decisión está determinada por sólo un pequeño conjunto de ejemplos de entrenamiento, llamados Support Vectors. (Sebastiani, 2002)

Figura N° 31: Aprendizaje del Support Vector Classifiers



Fuente: Machine Learning in Automated Text Categorization (Sebastiani, 2002)

En la Figura N° 31, las pequeñas cruces y círculos representan ejemplos de entrenamiento positivo y negativo, respectivamente, mientras que las líneas representan superficies de decisión. La superficie de decisión  $\sigma_i$  (indicada por la línea más gruesa) es, entre las que se muestran, la mejor posible, ya que es el elemento medio del conjunto más amplio de superficies de decisión paralelas (es decir, su distancia mínima a cualquier ejemplo de entrenamiento es máxima). Los cuadrados pequeños indican los vectores de soporte. (Sebastiani, 2002)

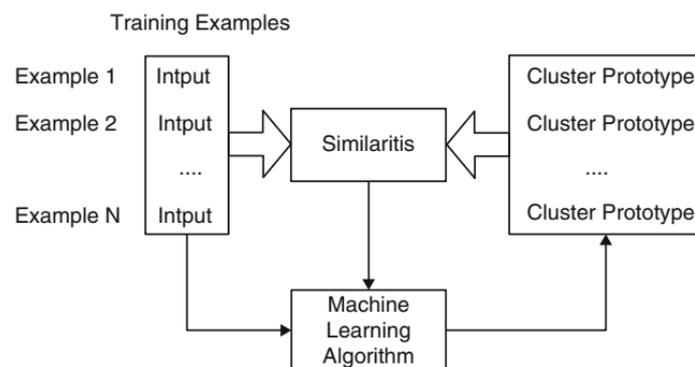
#### 2.2.4.1.2. Aprendizaje no supervisado

Se denomina aprendizaje no supervisado, al proceso de aprendizaje que depende de las similitudes de los datos de entrada para la tarea de clustering. Podemos mencionar los algoritmos de aprendizaje no

supervisados, como: el algoritmo de k-means y las redes de Kohonen, como enfoques para la clasificación de texto. (Jo, 2019)

Como se puede observar en la Figura N° 32, los prototipos de clústeres se inicializan al azar y se calculan sus similitudes con los ejemplos de entrenamiento. Los prototipos de clústeres están optimizados para maximizar sus probabilidades de obtener ejemplos de entrenamiento, a este proceso de aprendizaje se le conoce como no supervisado.

Figura N° 32: Aprendizaje No Supervisado



Fuente: Text Mining Concepts, Implementation, and Big Data Challenge (Jo, 2019)

Se asume que los ejemplos sin etiquetar son dados como ejemplos de entrenamiento, la colección no es particionado en el conjunto de entrenamiento y el conjunto de prueba como los algoritmos de aprendizaje supervisados. En este caso cada cluster es inicialmente caracterizado por sus propios prototipos y los ejemplos dados se organizados, en función de sus similitudes con los prototipos del cluster, cada ejemplo pertenece al cluster cuya similitud es mayor. (Jo, 2019)

### **2.2.5. Minería de texto**

El término "minería de texto" se usa generalmente para denotar todas las tareas necesarias para analizar grandes cantidades de texto en lenguaje natural y detectar patrones de uso léxico o lingüístico en un intento de extraer información probablemente útil (aunque solo probablemente correcta). (Sebastiani, 2002)

La minería de texto es el proceso de extraer el conocimiento implícito de los datos textuales. Debido a que el conocimiento implícito que es el resultado de la minería de texto no existe en el almacenamiento dado, debe distinguirse de la información que se recupera del almacenamiento. La clasificación, agrupación y asociación de texto son las tareas típicas de la minería de texto. (Jo, 2019)

En la minería de texto, la clasificación se refiere a la actividad de analizar o estudiar la colección de documentos de texto preclasificados para derivar un modelo o función que se puede usar para agrupar otros documentos de clase desconocidos en una o más clases predefinidas. (Kurnia & Suharjito, 2018)

#### **2.2.5.1. Minería de texto y minería de datos**

Así como la minería de datos se puede definir como la búsqueda de patrones en los datos, la minería de texto se trata de la búsqueda de patrones documentos de texto. Sin embargo, la similitud superficial entre los dos esconde diferencias reales.

Con la minería de texto, la información que se va a extraer se indica clara y explícitamente en el texto. No está oculto en absoluto. La minería de datos se puede caracterizar como la extracción de información implícita, previamente desconocida y potencialmente útil de los datos. (Witten & Frank, Data Mining Practical Machine Learning Tools and Techniques Second Edition, 2005)

Desde un punto de vista humano, el único sentido en el que es “previamente desconocido” es que las restricciones de tiempo hacen inviable que las personas lean el texto por sí mismas. El problema, por supuesto, es que la información no está redactada de una manera que sea susceptible de procesamiento automático. La minería de texto se esfuerza por presentarlo en una forma adecuada para el consumo de computadoras o de personas que no tienen tiempo para leer el texto completo. (Witten & Frank, Data Mining Practical Machine Learning Tools and Techniques Second Edition, 2005)

Un requisito común que deben cumplir tanto la minería de datos y la minería de texto es que la información que se extrae sea probablemente útil para los usuarios finales, esto significa que se pueda procesar y además sea capaz de brindar una base para que las decisiones basadas en datos se tomen de forma automática.

En el caso de la minería de datos, se puede expresar de una manera relativamente independiente del dominio: los patrones procesables son

aquellos que permiten realizar predicciones sobre nuevos datos de la misma fuente. El rendimiento se puede medir contando los éxitos y los fracasos, además se pueden aplicar técnicas estadísticas para comparar diferentes métodos de minería de datos sobre el mismo problema. Sin embargo, en la minería de texto es difícil caracterizar lo que significa "procesable" de una manera que sea independiente del dominio particular. Esto hace que sea difícil encontrar medidas de éxito justas y objetivas. (Witten & Frank, Data Mining Practical Machine Learning Tools and Techniques Second Edition, 2005)

En muchas aplicaciones de minería de datos, "potencialmente útil" recibe una interpretación diferente: la clave del éxito es que la información extraída debe ser comprensible en el sentido de que ayuda a explicar los datos. Esto es necesario siempre que el resultado esté destinado al consumo humano en lugar de (o además de) una base para una acción automática. Este criterio es menos aplicable a la minería de texto porque, a diferencia de la minería de datos, la entrada en sí es comprensible. La minería de texto con resultados comprensibles equivale a resumir características destacadas de un gran cuerpo de texto, que es un subcampo por derecho propio: resumen de texto. (Witten, Text mining, 2005)

#### **2.2.5.2. Clasificación de texto**

La clasificación o categorización de texto es el proceso de asignar un documento en una categoría o clase de acuerdo con las características del documento de entrada. (Kurnia & Suharjito, 2018)

Adicionalmente, según (Jo, 2019) la clasificación de texto se define como el proceso de asignar a cada elemento de datos a una categoría o algunas de las categorías de entre las predefinidas. Como tareas preliminares para la clasificación se tienen las siguientes: construir una lista de categorías predefinida, así como el sistema de clasificación con el cual los elementos de datos se asignarán a cada categoría. Para la tarea de clasificación, se considera dos tipos de enfoques:

- Basado en Reglas: donde las reglas simbólicas se definen manualmente y cada elemento de datos se clasifica por las reglas.
- Basados en Machine Learning: donde la capacidad de clasificación se construye con los datos de muestra y cada elemento es clasificado por él.

La categorización de texto es la tarea de asignar un valor booleano a cada par  $\langle d_j, c_i \rangle \in D \times C$  donde  $D$  es el dominio de documentos y  $C = \{c_1, \dots, c_{|C|}\}$  es un conjunto predefinido de categorías. Un valor T asignado a  $\langle d_j, c_i \rangle$  indica la decisión de clasificar  $d_j$  bajo  $c_i$ , mientras que un valor F indica una decisión de no clasificar  $d_j$  bajo  $c_i$ . Más formalmente, la tarea es aproximar la función objetivo desconocida  $\check{\Phi} : D \times C \rightarrow \{T, F\}$

(que describe cómo deben clasificarse los documentos) mediante una función  $\Phi : D \times C \rightarrow \{T, F\}$  llamado clasificador (también conocido como regla, hipótesis o modelo) tal que  $\check{\Phi}$  y  $\Phi$  "coincidan tanto como sea posible".

Cómo definir y medir con precisión esta coincidencia conocida como efectividad. (Sebastiani, 2002)

La evaluación de los clasificadores de texto normalmente se lleva a cabo de forma experimental, en lugar de analítica. La razón es que para evaluar analíticamente un sistema (por ejemplo, probar que el sistema es correcto y completo), necesitaríamos una especificación formal del problema que el sistema está tratando de resolver (por ejemplo, con respecto a qué precisión y exhaustividad son definidos), y la noción central de la clasificación de texto (es decir, la pertenencia de un documento a una categoría) esto, debido a su carácter subjetivo, inherentemente no formalizable. (Sebastiani, 2002)

Por lo tanto, a partir de ahora asumiremos que:

- Las categorías son solo etiquetas simbólicas, y no se dispone de ningún conocimiento adicional (de carácter procedimental o declarativo) de su significado.
- No se dispone de conocimientos exógenos, es decir, los datos proporcionados con fines de clasificación por una fuente externa; por lo tanto, la clasificación debe realizarse sobre la base del conocimiento endógeno únicamente, es decir, el conocimiento extraído desde los documentos. En particular, esto significa que metadatos como, por ejemplo: la fecha de publicación, el tipo de documento, la fuente de publicación, etc., se asume que no esté disponible.

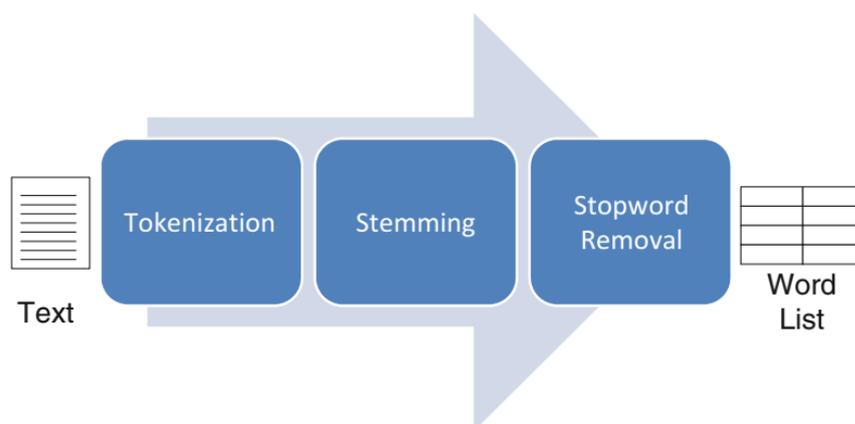
### 2.2.5.2.1. Indexación de texto

La indexación de texto se define como el proceso de convertir un texto o lista de textos en una lista de palabras, esto debido a que los textos se dan como formas no estructuradas, es casi imposible procesarlos directamente en su forma original mediante el uso de un programa de computadora. (Jo, 2019)

La indexación de texto consiste en segmentar un texto que consta de sentencias en palabras individuales. Una lista de palabras es el resultado de indexar un texto que luego se convertirá en la entrada para el proceso de clasificación.

Según la investigación de (Jo, 2019), los tres pasos básicos que se debe tener en cuenta para la construcción de un proceso de la indexación de texto se muestran en la Figura N° 33.

Figura N° 33: Los 3 pasos básicos en la Indexación de Texto



Fuente: Text Mining Concepts, Implementation, and Big Data Challenge (Jo, 2019)

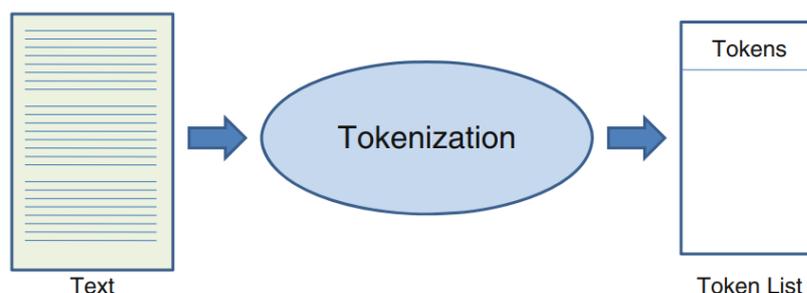
### 2.2.5.2.1.1. Tokenización

Podemos definir a la tokenización como el proceso de segmentar un documento separándolos en tokens o palabras a través del carácter espacio en blanco o signos de puntuación. Se puede realizar la tokenización a través de programas de computadora escritos en lenguajes de programación como: C, C++, Java, Python, etc. así como a los textos escritos en lenguaje natural, pero existe también la posibilidad de realizar un análisis morfológico. Sin embargo, el alcance de esta investigación está restringido solo a texto.

Para poder tokenizar textos escritos en lenguajes orientales tales como: chino, japonés y coreano es necesario realizar un análisis morfológico. (Jo, 2019)

En la Figura N° 34, podemos visualizar el proceso de tokenización en donde el ingreso del proceso es un texto y la salida del proceso es una lista de tokens. Durante el proceso de tokenización, el texto de ingreso ha sido segmentado en tokens a través de espacios en blanco o los signos de puntuación.

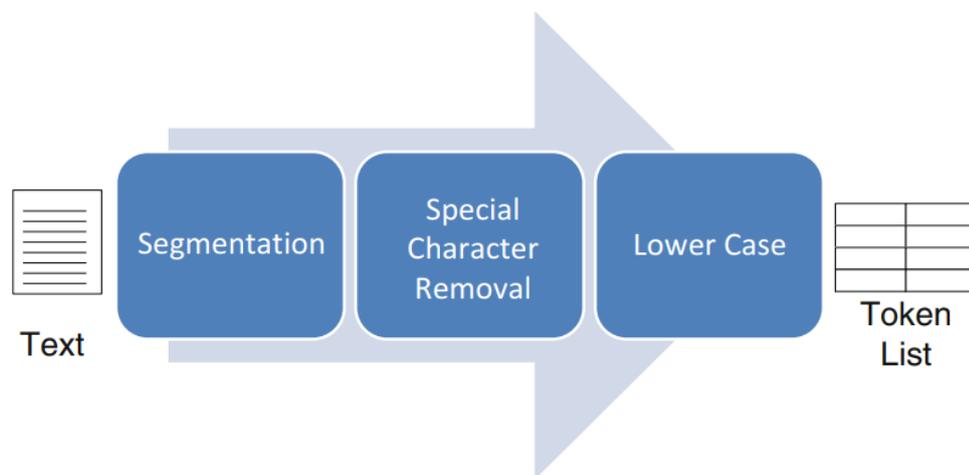
Figura N° 34: Proceso de Tokenización



Fuente: Text Mining Concepts, Implementation, and Big Data Challenge (Jo, 2019)

Como se muestra en la Figura N° 35, finalmente se eliminan las palabras que incluyen caracteres especiales o valores numéricos, y los tokens son transformados a sus caracteres en minúsculas. La lista de tokens se convierte en la entrada de los siguientes pasos de indexación de texto: Stemming o Stop-Word Removal.

Figura N° 35: Proceso de Tokenización de texto

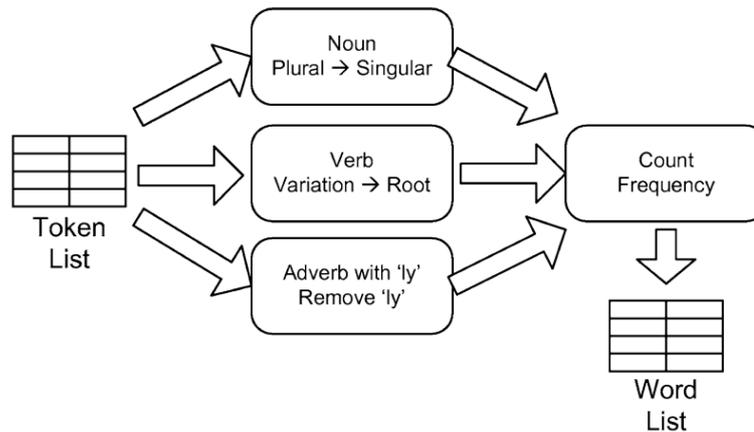


Fuente: Text Mining Concepts, Implementation, and Big Data Challenge (Jo, 2019)

#### 2.2.5.2.1.2. Stemming

Es el proceso de convertir cada token en su propia forma raíz utilizando reglas gramaticales. De acuerdo con (Jo, 2019) , Stemming es el proceso de mapear cada token generado en el proceso anterior de Tokenización en su propia forma raíz. Las reglas del Stemming, que son las reglas de asociación de tokens con su propia forma de raíz, son necesarias para su implementación. Stemming usualmente se son aplicados a sustantivos, verbos y adjetivos, como se muestra en la Figura N° 36. La lista de formas raíz se genera como resultado de este proceso.

Figura N° 36: Proceso de Stemming



Fuente: Text Mining Concepts, Implementation, and Big Data Challenge (Jo, 2019)

#### 2.2.5.2.1.3. Stop-word removal

Es el proceso de eliminación de palabras vacías o gramaticales como artículos, conjunciones y preposiciones de la lista de tokens.

Las palabras vacías son las palabras gramaticales que son irrelevantes para el contenido del texto, por lo que deben eliminarse para una mayor eficiencia. La lista de palabras vacías se carga desde un archivo.

La palabra vacía se refiere a la palabra que funciona solo gramaticalmente y es irrelevante para el contenido del texto dado. Preposiciones, como "en", "sobre", "a" y así sucesivamente, normalmente pertenecen al grupo de palabras vacías. Conjunciones como "y", "o", "pero" y "sin embargo" también pertenecen al grupo. El artículo definido, "el", y los artículos infinitos, "un" y "una", también son palabras vacías más frecuentes. Las palabras vacías aparecen predominantemente en todos los textos de la colección; eliminarlos

hace que mejore mucho la eficiencia en el procesamiento de textos. (Jo, 2019)

#### **2.2.5.2.1.4. Term weighting**

Term Weighting se refiere al proceso de calcular y asignar el peso de cada palabra según su grado de importancia.

El problema que nos enfrentamos en la clasificación de texto es que los datos están en forma de texto puro, por lo que este texto debe transformarse en un vector. La transformación de un texto en un vector es conocido como vectorización de texto. Es un proceso fundamental que se debemos seguir porque ningún algoritmo de Machine Learning entiende texto,

Según (Jo, 2019), TF-IDF (Term Frequency-Inverse Document Frequency) es el esquema más popular de ponderación de palabras en el área de recuperación de información, pero requiere las referencias a la colección de texto completa que se denomina corpus. El peso de las palabras se calcula mediante el esquema TF-IDF las cuales son proporcionales a las ocurrencias en el texto dado, pero inversamente proporcionales a las de otros textos. El peso TF-IDF,  $w_{ij}$ , de la palabra,  $t_i$ , en el texto,  $d_i$ , se calcula mediante la Ecuación 1 :

$$W_{ij} = \log \frac{N}{DF_i} (1 + \log TF_i) \quad (1)$$

Donde:

- N es el número total de textos de la colección

- $DF_i$  es el número de textos que incluyen la palabra  $t_i$  en la colección
- $TF_i$  es la ocurrencia de la palabra  $t_i$  en el texto  $d_i$ .
- Si la palabra aparece una sola vez en el texto, su peso depende del número de textos que la incluyan en el corpus, suponiendo que  $N$  sea una constante.
- Al codificar un texto en un vector numérico, se asigna cero a la palabra que no aparece en absoluto en el texto.

### **2.2.5.3. Aplicaciones de clasificación de texto**

La clasificación de texto se remonta al trabajo fundamental de (Maron, 1961) quien desarrolla su investigación la noción de indexación probabilística automática. Una suposición implícita detrás de su investigación ha sido afirmar que un documento pertenece a una categoría temática determinada o no, que no existe un término medio en la indexación.

Según (Sebastiani, 2002), desde entonces la clasificación de texto, se ha utilizado para una serie de aplicaciones diferentes, de las cuales aquí se revisarán brevemente las más importantes. Se debe tener en cuenta que los límites entre las diferentes clases de aplicaciones enumeradas aquí son difusos y de alguna manera artificiales, y algunos de estos pueden ser consideradas casos especiales de otros. Otras aplicaciones de clasificación de texto y que no se discutirá en la presente investigación son:

- Categorización del habla mediante una combinación de reconocimiento de voz y clasificación de texto.

- Categorización de documentos multimedia mediante el análisis de leyendas textuales.
- Identificación del autor para textos literarios de autoría desconocida o en disputa.
- Identificación del idioma para textos de idioma desconocido.
- Identificación automatizada del género del texto.
- Calificación de ensayos automatizada.

#### **2.2.5.4. Enfoque basado en reglas**

Según (Sebastiani, 2002), en los años 80, el enfoque más popular para la creación de clasificadores automáticos de documentos consistía en construir manualmente, mediante técnicas de KE (Knowledge Engineering), un sistema experto capaz de tomar decisiones de categorización de texto. Un sistema experto de este tipo consistiría típicamente en un conjunto de reglas lógicas definidas manualmente, una por categoría, de tipo

**if <DNF formula> then <categoría>**

Una fórmula DNF (disjunctive normal form) es una disyunción de cláusulas conjuntivas, el documento se clasifica en una <categoría> si satisface la fórmula, es decir, si satisface al menos una de las cláusulas. Un ejemplo de este enfoque se puede visualizar en la Figura N° 37:

Figura N° 37: Clasificador Basado en Reglas para la categoría de Trigo

<b>if</b>	<i>((wheat &amp; farm)</i>	<b>or</b>
	<i>(wheat &amp; commodity)</i>	<b>or</b>
	<i>(bushels &amp; export)</i>	<b>or</b>
	<i>(wheat &amp; tonnes)</i>	<b>or</b>
	<i>(wheat &amp; winter &amp; <math>\neg</math> soft))</i>	<b>then</b> WHEAT <b>else</b> $\neg$ WHEAT

Fuente: Machine Learning in Automated Text Categorization

El inconveniente de este enfoque es el cuello de botella en la adquisición de conocimientos. Es decir, las reglas deben ser definidas manualmente por un ingeniero del conocimiento con la ayuda de un experto en el dominio, si el conjunto de categorías se actualiza, entonces estos dos profesionales deben intervenir nuevamente, y si el clasificador se traslada a un dominio completamente diferente, un experto de dominio diferente debe intervenir y el trabajo debe repetirse desde cero. (Sebastiani, 2002)

#### 2.2.5.5. Enfoque basado en Machine Learning

El enfoque de Machine Learning (ML) para la Clasificación de Texto (CT), es un proceso inductivo general, el cual construye automáticamente un clasificador para una categoría  $ci$  al observar las características de un conjunto de documentos clasificados manualmente en  $ci$  o  $\bar{c}_i$  por un experto en el dominio, a partir de estas características, el proceso inductivo recoge las características que debe tener un nuevo documento invisible para ser clasificado en el  $ci$ . (Sebastiani, 2002)

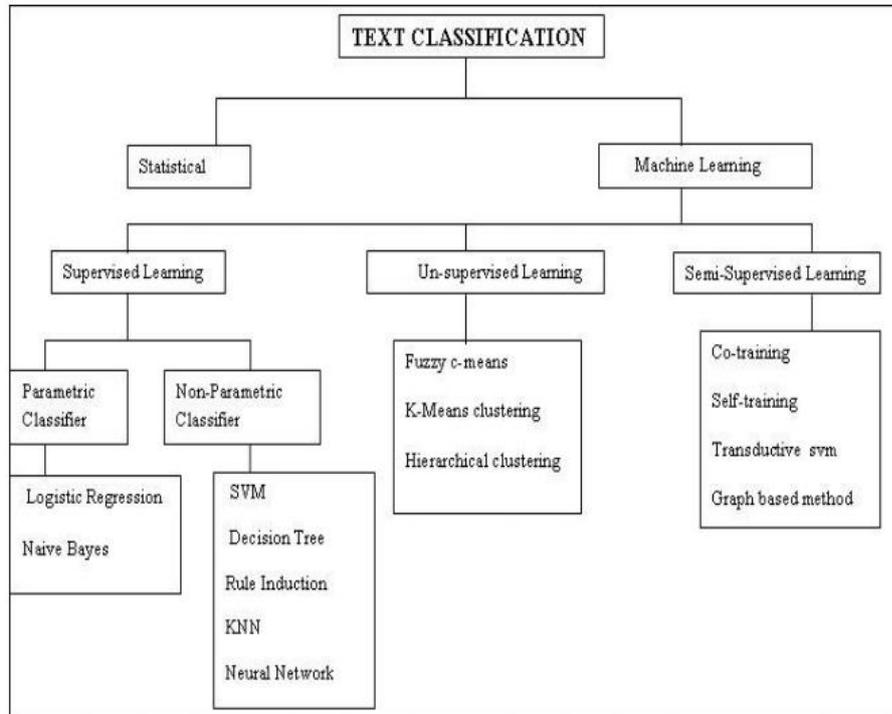
En terminología de Machine Learning (ML), el problema de clasificación es una actividad de aprendizaje supervisado, ya que el proceso de aprendizaje

es “supervisado” por el conocimiento de las categorías y de las instancias de entrenamiento que les pertenecen. (Sebastiani, 2002)

Las ventajas del enfoque de Machine Learning (ML) sobre el enfoque basado en reglas (KE) son evidentes. El esfuerzo de ingeniería se dirige hacia la construcción no de un clasificador, sino de un constructor automático de clasificadores (learner). Esto significa que, si un learner está disponible, todo lo que se necesita es la construcción inductiva y automática de un clasificador a partir de un conjunto de documentos clasificados manualmente. Lo mismo sucede si ya existe un clasificador y el conjunto original de categorías se actualiza, o si el clasificador se traslada a un dominio completamente diferente. (Sebastiani, 2002)

Según (Thangaraj & Sivakami, 2018), las técnicas de Machine Learning (ML) se inventaron especialmente para la automatización, en la Figura N° 38, podemos visualizar que los algoritmos de clasificación se dividen ampliamente en categorías supervisadas, no supervisadas y semi supervisadas de acuerdo con los criterios de aprendizaje seguidos.

Figura N° 38: Representación de Clasificadores de Texto



Fuente: Text Classification Techniques: A Literature Review (Thangaraj & Sivakami, 2018)

### 2.2.5.5.1. Conjunto de entrenamiento

El enfoque de Machine Learning se basa en la disponibilidad de un corpus inicial  $\Omega = \{d_1, \dots, d_{|\Omega|}\} \subset D$  de documentos preclasificados bajo  $C = \{c_1, \dots, c_{|C|}\}$ . Es decir, los valores de la función total  $\Phi : D \times C \rightarrow \{T, F\}$  son conocidos por cada par  $\langle d_j, c_i \rangle \in \Omega \times C$ . Un documento  $d_j$  es un ejemplo positivo de  $c_i$  si  $\Phi(d_j, c_i) = T$ , un ejemplo negativo de  $c_i$  si  $\Phi(d_j, c_i) = F$ .

En los entornos de investigación y también en la mayoría de los entornos operativos, una vez que se ha construido un clasificador  $\Phi$ , es conveniente evaluar su precisión. En este caso, antes de la construcción del clasificador,

el corpus inicial se divide en dos conjuntos, no necesariamente del mismo tamaño:

#### **2.2.5.6. Análisis de sentimiento**

El análisis de sentimiento es también conocido como minería de opinión, el cual analiza la opinión y las emociones que expresan las personas hacia entidades como productos, organizaciones y sus atributos asociados. (Tripathy, Agrawal, & Rath, 2016)

El análisis de sentimiento se refiere al proceso de clasificar una opinión en positiva, negativa o neutro. Una opinión es un dato textual de entrada y una de las tres actitudes se genera como salida. (Jo, 2019)

El análisis de sentimiento es un campo de estudio que analiza las opiniones, sentimientos, evaluaciones, valoraciones, actitudes y emociones de las personas hacia las organizaciones. (Alalwan, Rana, Dwived, & Algharabat, 2017)

El análisis de sentimiento se utiliza también para clasificar automáticamente las opiniones sobre productos comerciales o temas políticos en función a las actitudes de las personas. (Jo, 2019)

De acuerdo con los autores (Alalwan, Rana, Dwived, & Algharabat, 2017) con frecuencia, el sentimiento se considera una creencia o juicio personal que no se basa en el razonamiento racional, sino en la emoción subjetiva.

El análisis de sentimiento se lleva a cabo en tres niveles diferentes, como son el nivel de documento, nivel de oración y nivel de aspecto. El nivel de documento clasifica si la opinión del documento es positiva, negativa o neutral. El nivel de la oración determina si la oración expresa alguna opinión negativa, positiva o neutral. El nivel de aspecto se centra en todas las expresiones de sentimientos presentes en un documento dado y el aspecto al que se refiere. (Tripathy, Agrawal, & Rath, 2016)

El análisis de sentimiento se puede aplicar al análisis de sentimiento del texto, el análisis de sentimiento del habla y el análisis de sentimiento visual. Aunque la tarea es difícil solo con declaraciones verbales. (Mohan Debarchan Mohanty & Mihir Narayan Mohanty, *Advanced Data Mining Tools and Methods for Social Computing*, 2022)

Un procedimiento de Minería de texto tiene que referirse, implícita o explícitamente, a un enfoque teórico sociológico o psicológico, que explica la producción del lenguaje y la interacción social que establece el intercambio de comunicación.

Para ser riguroso, un estudio debe hacer coincidir el enfoque teórico con el metodológico, pero, sorprendentemente, este aspecto es aparentemente descuidado por los investigadores (Alalwan, Rana, Dwived, & Algharabat, 2017).

La mayoría de la literatura sobre el procedimiento de Minería de Texto llama la atención sobre la metodología (etiquetado de palabras, estructura léxica

de la oración, procedimiento estadístico, etc.), centrándose en el contenido manifiesto del texto.

La mayoría de los métodos se basan en un enfoque de arriba hacia abajo en el que se realiza un procedimiento de codificación a priori de términos, o texto, centrado en el contenido manifiesto de la palabra.

Siguiendo un enfoque de arriba hacia abajo, estos métodos utilizan categorías de contenido predefinidas para clasificar semánticamente el texto. (Balbi, Misuraca, & Scepi, 2018)

Cada una de estas categorías corresponde a un diccionario temático que contiene todas las palabras indicativas del contenido representado por esa categoría. (Greco & Polli, 2019)

Sin embargo, como lo destacó Saussure en un Curso de Lingüística General, el lenguaje es un sistema de signos que expresa un sistema de significado.

A pesar de que los enfoques arriba-abajo de la Minería de Texto permiten una investigación confiable y válida, presentan una limitación importante, sin tener en cuenta la naturaleza contextual del significado lingüístico

Por lo tanto, un término puede asumir un significado específico de acuerdo con su asociación con los otros términos del texto. (Carli, R. & Paniccia, R. M., 2004)

Según lo establecido por Liu (2015), no es suficiente clasificar el léxico del sentimiento para realizar un análisis del sentimiento porque un término, clasificado como una palabra del sentimiento positivo o negativo, puede tener una orientación opuesta según el contexto.

De hecho, el significado de una palabra es polisémico y está sujeto a la forma en que se combina con otras palabras en una interacción comunicativa, es decir, depende de su asociación con otras palabras.

Por ejemplo, "bomba" generalmente indica un sentimiento negativo, por ejemplo, "Hubo otra explosión de camión bomba esta mañana en el mercado en Sadr City", pero también puede implicar un sentimiento positivo de admiración, por ejemplo, "¡ella es una bomba sexual!".

Por lo tanto, la presencia o ausencia de palabras de sentimiento en una oración no implica necesariamente la posibilidad de clasificar un sentimiento.

Es decir, una oración que contiene términos de sentimiento puede ser neutral, lo que ocurre con frecuencia en preguntas u oraciones condicionales, y una oración sin una palabra de sentimiento puede expresar una opinión. Además, las oraciones sarcásticas, con o sin palabras de sentimiento, son difíciles de clasificar. (Liu, 2015)

En base a estas consideraciones, Minería de Texto Emocional (ETM) según Greco (2016) es un procedimiento de Minería de Texto que, mediante su lógica abajo-arriba, permite un enfoque de Minería de Texto sensible al

contexto en datos no estructurados, que constituye el 95% de Big Data (Gandomi & Haider, 2015).

ETM es un procedimiento de minería de texto no supervisado, basado en un enfoque socio constructivista y un modelo psicodinámico. De acuerdo con este enfoque, el sentimiento no es solo la expresión de un estado de ánimo, sino también la evidencia de un proceso de pensamiento social y latente que establece las interacciones, el comportamiento, las expectativas y la comunicación de las personas. (Greco & Polli, 2019)

Sabemos que el comportamiento de una persona depende no solo de su razonamiento racional sino también, y, sobre todo, de su modo de funcionamiento mental y emocional. En otras palabras, las personas categorizan conscientemente la realidad y, al mismo tiempo, la simbolizan inconscientemente emocionalmente, para adaptarse a su entorno social (Greco & Polli, 2019)

La categorización consciente y la simbolización inconsciente son dos procesos mentales paralelos que siguen dos reglas de funcionamiento diferentes, es decir, dos lógicas. (Greco & Polli, 2019)

La simbolización inconsciente es social, ya que las personas la generan interactivamente y comparten los mismos significados emocionales a través de esta interacción (Greco F. , 2016).

Dado que la comunicación y el comportamiento son el resultado de este funcionamiento social mental, es posible analizar la comunicación (texto) para inferir el funcionamiento social mental (matriz simbólica) y explicar, o predecir, el comportamiento de las personas en diferentes contextos. Además, explicar o pronosticar el comportamiento de las personas a través de su comunicación en las redes sociales es relevante para la gestión empresarial. (Gloor, 2017)

Debido al hecho de que el proceso consciente establece el contenido manifiesto de la comunicación, es decir, lo que se comunica, el proceso inconsciente se puede inferir a través de cómo se comunica, es decir, las palabras elegidas para comunicarse y su asociación dentro del texto.

Consideramos que las personas simbolizan emocionalmente un evento o un objeto, y comparten socialmente esta simbolización. Las palabras que eligen para discutir un evento u objeto son el producto del funcionamiento mental inconsciente socialmente compartido. (Greco F. , 2016)

#### **2.2.5.7. Minería de texto en redes sociales**

El incremento de dispositivos con acceso a Internet, así como el aumento de usuarios en las plataformas de redes sociales ha hecho que exista una gran cantidad de datos de texto no estructurados.

La Minería de Texto es una extensión de las técnicas de Minería de Datos y se basa en representaciones de documentos basadas en características que

pueden incluir palabras, conceptos, caracteres y términos. (Nishanth, Ravi, Ankaiah, & Bose, 2012)

Las técnicas de Minería de Texto incluyen los procesos de edición, organización y análisis de una gran cantidad de documentos para proporcionar a los investigadores información específica e identificar las relaciones identificadas por la información. (Sullivan, D., 2001)

Muchas empresas han adoptado la Minería de las Redes Sociales, como el uso de herramientas de medios sociales para proporcionar diversos servicios e interactuar con los clientes. Esto puede ayudar a las empresas a comprender cómo realizar un análisis competitivo y transformar los datos en conocimiento para la toma de decisiones.

Esta técnica efectiva se utiliza para extraer el valor comercial de la gran cantidad de datos de redes sociales disponibles y puede proporcionar ayuda a las empresas para desarrollar su estrategia de análisis competitivo de redes sociales. (He, Zha, & Li, 2013)

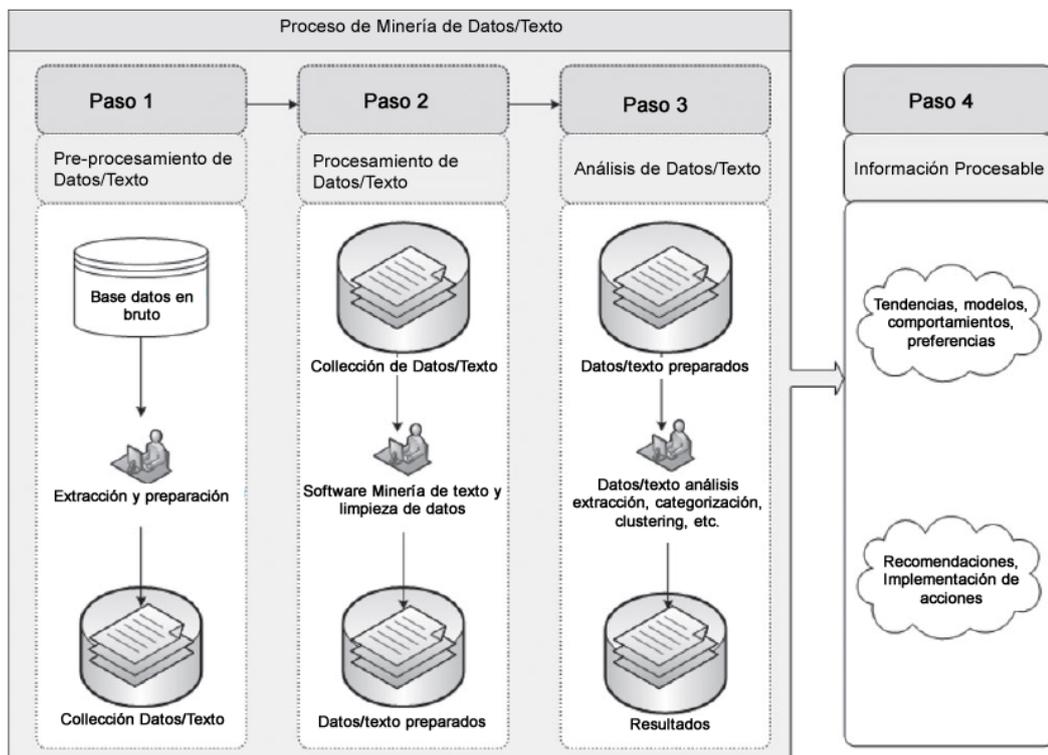
El proceso de Minería de Texto incluye varios pasos. Primero se hace el preprocesamiento de los datos, en el cual los datos sin procesar se transformaron en un formato utilizable.

Principalmente mediante la limpieza, la asignación de atributos y la integración de datos. Posteriormente, se aplican varias técnicas de extracción de datos y extracción de texto para examinar los conjuntos de

datos con el fin de obtener información sobre las actividades de los usuarios en las redes sociales. (He, Zha, & Li, 2013)

Para entender de una manera gráfica los pasos necesarios que se tienen que ejecutar en un proceso de Minería de Texto, los autores Abdous y He (2011) proponen el siguiente modelo el cual se muestra en la Figura N° 39

Figura N° 39: Proceso de Minería de Texto



Fuente: Using text mining to uncover students' technology-related problems in live video streaming (Abdous & He, 2011)

La Minería de Texto implica combinar un documento o recurso de texto para obtener información estructurada valiosa.

Esto requiere herramientas analíticas sofisticadas que procesen texto para recoger palabras clave específicas o puntos de datos clave de lo que se consideran formatos relativamente sin formato o no estructurados.

En Minería de Texto los sistemas utilizan las taxonomías y análisis léxico para determinar qué partes de un documento de texto son valiosas como datos minados.

Los modelos estadísticos suelen ser útiles, y los sistemas también pueden usar heurísticas, o conjeturas algorítmicas, para tratar de determinar qué partes de un texto son importantes.

Otros sistemas de control incluyen etiquetado y análisis de palabras clave, donde las herramientas buscan nombres propios específicos u otras etiquetas y palabras clave para averiguar sobre qué se está escribiendo.

#### **2.2.6. Índice de promotores neto (NPS)**

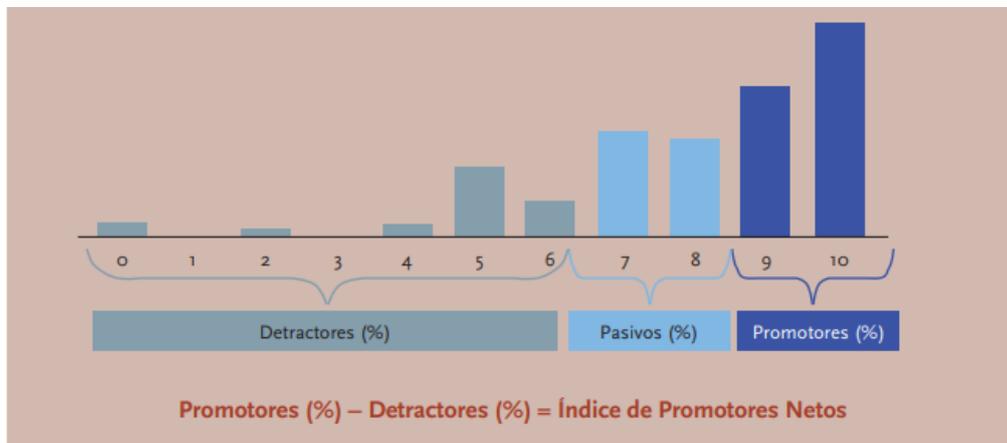
El Índice de Promotores Neto (NPS) es una medida que se utiliza en el marketing para evaluar la lealtad del cliente hacia una empresa específica. El proceso comienza con una pregunta simple pero poderosa en una escala de 1 (nada probable) a 10 (extremadamente probable): "¿Qué tan probable es que recomiende nuestra empresa a un amigo o colega?" (Reichheld, 2003)

Como se puede observar en la Figura N° 40, es una forma cómoda y sencilla de analizar la satisfacción de los clientes. Los clientes califican su

satisfacción en una escala de 0 a 10 sobre una empresa, un servicio o un producto que utilizan. Los resultados se utilizan para segmentar a los clientes en tres grupos: promotores, pasivos y detractores. Luego, el porcentaje de detractores se resta del porcentaje de promotores para obtener una puntuación de promotor neto. (Jouve, Martin, & Guerin, 2012)

- **Detractores** se derivan de clientes con satisfacción del cliente entre 0 y 6. Son clientes insatisfechos. Es probable que muchos de ellos abandonen. Pueden inscribirse en programas de retención o acciones de seguimiento para solucionar sus problemas y convertirlos en clientes satisfechos. Sus comentarios se valoran para comprender mejor lo que va mal y comprender mejor los criterios clave asociados con la calidad del servicio.
- **Pasivos** corresponden a la satisfacción del cliente que es igual a 7 u 8. Se supone que son neutrales. Se supone que deben seguir siendo clientes activos siempre que no experimenten ningún problema importante.
- **Promotores** corresponden a la satisfacción del cliente que es igual a 9 o 10. Son clientes muy satisfechos que probablemente difundirán un boca a boca positivo a su alrededor. Las empresas pueden comprender mejor las razones de su satisfacción y replicarlas en diferentes geografías o mercados.

Figura N° 40: Índice de Promotores Netos de una empresa



Fuente: El compromiso de los clientes como objetivo estratégico (Arjonilla, 2009)

### 2.2.7. Interfaz de programación de aplicaciones (API)

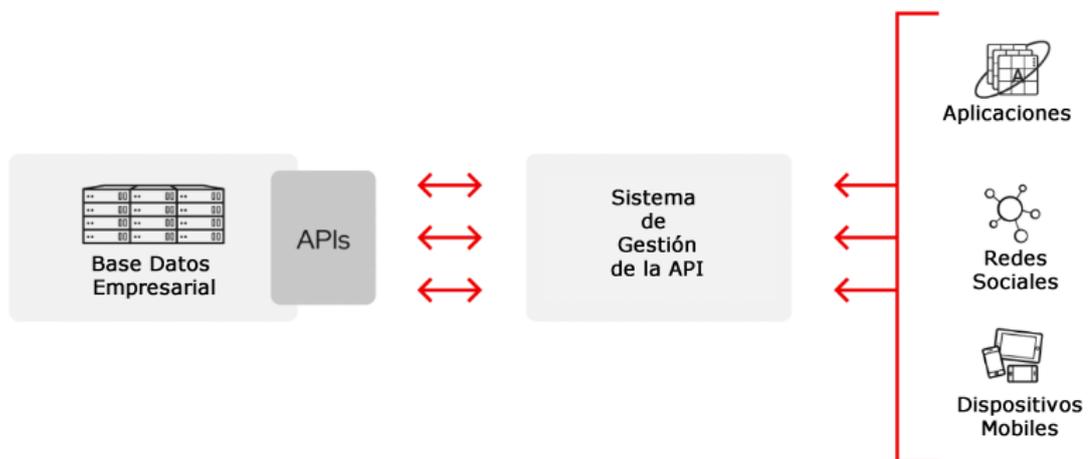
Es un conjunto de definiciones y protocolos que se utiliza para desarrollar e integrar el software de las aplicaciones. (RedHat, 2019)

Como se puede apreciar en la Figura N° 41, las API permiten que sus productos y servicios se comuniquen con otros, sin necesidad de saber cómo están implementados. Esto simplifica el desarrollo de aplicaciones y permite ahorrar tiempo y dinero. Las API le otorgan flexibilidad, simplifican el diseño, la administración y el uso de las aplicaciones, además proporcionan oportunidades de innovación, lo cual es ideal al momento de diseñar herramientas y nuevos productos.

Debido a que simplifican la forma en que los desarrolladores integran los elementos de aplicaciones nuevas en una arquitectura actual, las API permiten la colaboración entre el equipo comercial y el equipo de TI. Las necesidades comerciales cambian rápidamente en respuesta a los mercados

en constante cambio, donde la competencia puede modificar un sector entero con una aplicación nueva. Para seguir siendo competitivos, es importante admitir la implementación y el desarrollo rápidos de servicios innovadores.

Figura N° 41: Estructura de una API



Fuente: Redhat.com (2019)

Las API simplifican y amplían la forma en que las empresas pueden conectarse a aplicaciones externas, redes sociales, dispositivos móviles, pasarela de pagos, etc., permitiendo intercambiar información fácilmente.

### 2.2.8. RESTful API

REST, un acrónimo de Representational State Transfer, es un estilo de arquitectura de software que se utiliza para la construcción de aplicaciones web altamente escalables el cual fue definido por (Fielding, 2000) en su tesis doctoral.

Según (Prehofer & Gerostathopoulos, 2017), los principales ingredientes que debe tener un diseño RESTful son los siguientes:

- Identificación de recursos a través de Uniform Resource Identifier (URI). Estos están estructurados jerárquicamente y cada recurso debe tener al menos una URI.
- Interfaces uniformes para leer y manipular los recursos. Estas son las cuatro operaciones HTTP básicas: GET, POST, PUT y DELETE. Otras operaciones, por ejemplo. HEAD y OPTIONS, tratan con metadatos.
- Mensajes autodescriptivos. Se puede acceder a la representación de los recursos en diferentes formatos, por ejemplo: HTML, JSON o XML. Los mensajes, tanto de solicitud como de respuesta, contienen el contexto completo y son autodescriptivos en este sentido.
- Interacciones sin estado, es decir, el servidor no mantiene el estado de sesión en las interacciones con los clientes. Esto significa que toda la información para cumplir con una solicitud se incluye en la solicitud HTTP, es decir, el nombre del recurso y el mensaje.

El estilo REST proporciona un mecanismo uniforme para el acceso a los recursos que se encuentran disponibles en un servidor web, lo que simplifica el desarrollo de aplicaciones web. Su estructura garantiza un uso eficaz de Internet, en particular de intermediarios como cachés y proxies, lo que resulta en un acceso rápido a las aplicaciones. (Klein & Namjoshi, 2011)

Durante la última década, el uso de la arquitectura REST ha aumentado considerablemente y se ha convertido en el estándar de la industria para la implementación de proyectos de aplicaciones web a gran escala.

La otra cara de esto es una confusión considerable sobre los principios del diseño RESTful, que a menudo se malinterpretan y se aplican mal. Esto da como resultado aplicaciones que son funcionalmente correctas, pero que no logran todos los beneficios de flexibilidad y escalabilidad que son posibles con REST. (Klein & Namjoshi, 2011)

### **2.2.9. API Graph Facebook**

La API Graph es la forma principal de obtener datos de la plataforma de Facebook. Es una API basada en el protocolo HTTP que las aplicaciones pueden usar para consultar datos mediante un lenguaje de programación, así como realizar una amplia variedad de otras tareas.

En el mundo de las API, un "Explorador de API" es una interfaz que le ayuda a elaborar una solicitud a través de una URL. Esta URL es como una línea de comando que le dice a Facebook que haga algo en su nombre. Por ejemplo, realizar una solicitud GET para extraer datos de su perfil de Facebook.

#### **2.2.9.1. Estructura de la API Graph**

La API Graph es una representación de la información en Facebook. La estructura de la API Graph se compone de los siguientes elementos:

- **Nodos:** son objetos individuales, como un usuario, una foto, una página o un comentario.
- **Lados:** son conexiones entre una colección de objetos y un solo nodo, como fotos en una página o comentarios en una foto.
- **Campos:** son datos sobre un objeto que desea incluir en la respuesta, como el cumpleaños de un usuario o el nombre de una página.

Por lo general, se usan los nodos para obtener datos sobre un objeto específico, se usan los lados para obtener colecciones de objetos de un solo nodo y se usan los campos para obtener datos que se incluirán en la respuesta a la consulta, sobre un solo objeto o cada objeto de una colección.

### **2.3. MARCO CONCEPTUAL**

La Inteligencia de Negocios Social también conocido como la Inteligencia de Negocios aplicado a las Redes Sociales, es un paso adelante en la aplicación de la Inteligencia de Negocios porque permite obtener abundante información disponible de fuentes externas a los sistemas de soporte a las operaciones de negocio de las empresas.

Es un nuevo campo de aplicación en la Inteligencia de Negocios el cual permite extraer una gran cantidad de datos textuales los cuales se encuentran disponibles de las plataformas de redes sociales.

Aplicando técnicas de Minería de Texto categorizarlas y aplicando técnicas de Análisis de Sentimiento poder identificar formas de expresión para descubrir patrones de comportamiento de grupos de usuarios en redes sociales y a través de indicadores de gestión (KPI's) brindar información para que los directivos y gerentes de marca puedan tomar decisiones oportunas para diseñar estrategias de campañas de marketing o enviar mensajes claros a los consumidores en la plataforma de red social Facebook, acerca del lanzamiento de nuevos programas de seguros.

La Minería de Texto es el proceso de explorar y analizar grandes cantidades de datos de texto no estructurados ayudados por un software que puede identificar conceptos, patrones, temas, palabras clave y otros atributos en los datos. También se conoce como Análisis de Texto, aunque algunos investigadores hacen una distinción entre los dos términos, el Análisis de Texto es una aplicación que permite el uso de técnicas de Minería de Texto para clasificar un grupo de datos.

El Análisis de Sentimiento o Minería de Opinión analiza las distintas formas de expresión que tienen las personas cuando reaccionan de manera positiva, negativa o neutra ante una publicación o al escribir comentarios. Identificar o pronosticar el comportamiento que tienen los usuarios a través de comunicación en redes sociales es muy importante para las organizaciones para que puedan reaccionar rápidamente ante cualquier campaña de desprestigio de una marca o para diseñar estrategias de comunicación que permita llegar a los consumidores con un mensaje claro.

## **2.4. MARCO LEGAL**

En el Perú existe la Ley de Protección de Datos Personales – Ley N° 29733 el cual tiene como objetivo garantizar el derecho fundamental a la protección de datos personales. Se encuentra presente en el artículo N°2, numeral 6, de la Constitución Política del Perú.

Con esta ley, las empresas y entidades públicas están obligadas a garantizar la protección de los datos con los que cuentan en sus sistemas informáticos, con el fin de evitar que terceros no autorizados accedan a ellos. Solo están sujetos a la Ley aquellos datos personales que identifiquen o hacen identificable a una persona natural; para registros o información sobre personas jurídicas no se aplica la normativa.

La ley obliga a las entidades a gestionar los datos personales con el consentimiento del titular de los datos, recopilándose de manera expresa, verbal o física (en caso contengan datos sensibles). Y en caso se tratase de menores de edad, se debe realizar con el consentimiento de sus padres o tutores. Esta recopilación debe realizarse con relación a las finalidades determinadas, explícitas y lícitas. (Congreso de la República, 2011)

En el Perú no existe una ley que regule la prohibición del uso de información que será descargada de las redes sociales. De acuerdo a Facebook está información solo deberá ser usada para temas académicos.

Respecto a la privacidad de los datos en redes sociales uno de los principios que rige la legislación de protección de datos en todo el mundo es el

Principio de Finalidad el cual es recogido en el artículo 6 de la Ley de PDP que señala expresamente lo siguiente:

*Los datos personales deben ser recopilados para una finalidad determinada, explícita y lícita. El tratamiento de los datos personales no debe extenderse a otra finalidad que no haya sido la establecida de manera inequívoca como tal al momento de su recopilación, excluyendo los casos de actividades de valor histórico, estadístico o científico cuando se utilice un procedimiento de disociación o anonimización.*

Por este principio, los datos solo pueden ser utilizados para aquellas finalidades que el titular del banco de datos informó al titular de los mismos antes de recogerlos. Cualquier uso para una finalidad distinta constituye un tratamiento indebido y por lo tanto sujeto a la posibilidad de una sanción.  
(Gestion.com, 2017)

## **CAPÍTULO III**

### **METODOLOGÍA DE LA INVESTIGACIÓN**

#### **3.1. UNIDADES DE ANÁLISIS**

La unidad de análisis de la presente investigación es: “Comentarios en redes sociales de las empresas del sector asegurador de la ciudad de Lima Perú durante el año 2020”

Adicionalmente, para probar la hipótesis se ha utilizado la encuesta, la cual deberá ser respondida por los usuarios o encuestados del modelo de inteligencia de negocios basado en minería de opinión.

#### **3.2. MUESTRA**

##### **3.2.1. Muestra cuantitativa**

La muestra de la presente investigación serán los comentarios en Facebook de las empresas del sector asegurador que se encuentran en las 3 primeras posiciones del Ranking de Primas de Seguros Netas, realizadas por usuarios de 18 a 70 años, que pertenezcan a la comunidad y se conecten desde cualquier dispositivo y lugar.

El contenido será variado y estará compuesto de:

- Revisión de productos
- Opiniones
- Valoraciones
- Reacciones: positivo, negativo o neutro.

### **3.2.2. Muestra cualitativa**

Se llevará a cabo entrevistas a personas relacionadas a la industria de seguros cómo son: directivos, ejecutivos y analistas de negocios de tal manera de obtener el juicio de expertos para ajustar el insumo de entrada.

Asimismo, se realizarán encuestas a los usuarios de seguros para conocer el nivel de satisfacción que tienen respecto a los servicios entregados por las empresas del sector asegurador. Se considera que la información obtenida de estas encuestas será de gran utilidad para el desarrollo de la presente investigación.

La empresa en estudio cuenta con un total de 90 empleados administrativos de los cuales 20 empleados son ejecutivos de operaciones, que es el personal de servicio al cliente y el que utilizará el modelo de inteligencia negocios basado en minería de opinión.

### **3.3. PERIODO DE LOS DATOS**

Los datos extraídos para la presente investigación tendrán un periodo mensual.

### **3.4. TIPO Y NIVEL DE INVESTIGACIÓN**

#### **3.4.1. Tipo de investigación**

La presente investigación reúne las condiciones metodológicas de ser una investigación aplicada, debido a que se utilizarán diversos conocimientos relacionados al diseño de soluciones de inteligencia de negocios, minería de texto y análisis de sentimiento.

#### **3.4.2. Nivel de investigación**

De acuerdo con la naturaleza de la investigación, reúne por su nivel las características de ser: descriptivo y explicativo.

### **3.5. PERIODO DE ANÁLISIS**

El periodo de análisis de la presente investigación serán los comentarios publicados por usuarios en redes sociales desde el año 2020. La frecuencia de extracción será mensual.

### **3.6. FUENTES DE INFORMACIÓN E INSTRUMENTOS UTILIZADOS**

La fuente de información de la presenta investigación serán los comentarios de usuarios realizados a publicaciones de las empresas corredoras de seguros en la plataforma de redes sociales: Facebook.

### **3.7. TÉCNICAS E INSTRUMENTOS DE RECOLECCIÓN DE DATOS**

Para la presente investigación, la selección y elaboración de técnicas e instrumentos de recolección es parte importante dentro de la etapa de recolección de la información debido a que establece la ruta para encontrar la información requerida que dará respuesta al problema planteado.

### **3.7.1. Técnicas**

Las técnicas básicas para la recolección de la información es el medio mediante el cual el investigador se relaciona con los participantes que nos permita obtener la información necesaria que le permita alcanzar los objetivos planteados en la presente investigación.

La técnica seleccionada para la presente investigación es la Observación

#### **3.7.1.1. Minería de texto**

En proceso de análisis de contenido se utilizará la herramienta de clasificación de texto la cual permite ubicar el documento en una categoría o clase de acuerdo con las características del documento. En minería de texto la clasificación se refiere a la actividad de analizar o estudiar la colección de documentos de texto preclasificados para derivar un modelo o función que pueda usarse para agrupar otros documentos de clase desconocidas en una o más clases predefinidas en la presente investigación.

En la presente investigación los algoritmos de clasificación de texto utilizados son: Naive Bayes, Decision Tree y SVM. Cada algoritmo de clasificación de texto será evaluado para determinar su rendimiento. El método de evaluación que se utilizará es mediante la aplicación de un esquema de pruebas con 5 composiciones de datos diferentes y también para cada una de las composiciones de prueba se realizará aplicando el método de validación cruzada de 10 veces.

En la clasificación de texto hay varias formas de medir su desempeño, como el cálculo de: exactitud, precisión, recuperación. Luego, los resultados del procesamiento del texto se almacenarán en una tabla de la base de datos.

#### **3.7.1.2. Análisis de sentimiento**

El análisis de sentimiento o minería de opinión se refiere a la amplia área del procesamiento del lenguaje natural, la minería de textos, la lingüística computacional, que involucra el estudio computacional de sentimientos, opiniones y emociones expresadas en el texto. Aunque, la visión o actitud basada en la emoción en lugar de la razón se suele denominar un sentimiento. Por lo tanto, se debe asignar un equivalente para minería de opiniones o análisis de sentimientos.

#### **3.7.1.3. Encuestas**

La forma de obtener datos directamente de los usuarios del modelo de inteligencia de negocios propuesta en la presente investigación, son las encuestas a fin de conseguir sus opiniones o sugerencias respecto del modelo.

#### **3.7.2. Instrumentos**

Para la presente investigación se dispondrá de instrumentos que nos permitan la extracción de los comentarios que forman parte de la muestra, así como del procesamiento de estos para categorización en una clase del rubro de seguros.

### **3.7.2.1. API Graph**

Para el proceso extracción de comentarios se utilizará la herramienta: API Graph que es una herramienta expuesta por las plataformas de red social Facebook que permite a las aplicaciones leer información directamente desde las redes sociales. La API Graph es una herramienta fundamental para la presente investigación.

### **3.7.2.2. Facepager**

Facepager es una herramienta para la recuperación automatizada de datos públicos en la web, fue desarrollada por Jakob Jünger y Till Keyling. Con esta herramienta se puede descargar datos desde las redes sociales, sobre todo comentarios de usuarios, de YouTube, Twitter, Facebook y Amazon.

Facepager trae una serie de presets (configuración pre-establecida de consultas a los datos en Facebook) ya instalados en la herramienta. Además, los parámetros se pueden modificar y configurar para obtener datos más precisos y de mayor interés.

### **3.7.2.3. Cuestionarios**

Para obtener la información de retroalimentación de parte de los usuarios del modelo, se ha desarrollado cuestionarios utilizando una escala de Likert de tipo frecuencia, el cuestionario cuenta con (5) cinco opciones de tipo cerradas.

Siendo los valores: 4 y 5 los valores más altos, lo cual demuestran que los usuarios tienen mucho más interés por el modelo. Los valores: 1 y 2 son los

valores bajos lo cual demuestran que los usuarios no tienen interés en el uso del modelo inteligencia de negocios.

## **CAPÍTULO IV**

### **MARCO ORGANIZACIONAL**

#### **4.1. EMPRESA DE ESTUDIO**

La presente investigación comprende la propuesta de implementación de un proyecto de Inteligencia de Negocios para la toma de decisiones en la empresa Willis Corredores de Seguros S.A.

##### **4.1.1. Visión**

Ser considerados por nuestros clientes, y le mercado asegurador como el asesor analítico de riesgos más innovador y confiable para H&B, CRB Retail, Facultativo y Affinity.

##### **4.1.2. Misión**

Proveer la mejor asesoría en H&B, CRB Retail, Facultativo y Affinity. Estructurando e implementando soluciones Innovadoras que mitiguen riesgos, incrementen beneficios y promuevan el talento. Con el objetivo de Fortalecer el negocio y la fuerza laboral de nuestros clientes.

## **4.2. DESCRIPCIÓN DEL CONTEXTO**

Willis es una empresa con más de 180 años de experiencia en el mundo, desde 1828 Willis es reconocida por sus resultados e integridad. Es una compañía que siempre tiene como objetivo la calidad del servicio al cliente. Es por esto por lo que Willis construyó una cultura que es apasionada de lo que hace y siempre pone a los clientes en primer lugar.

Con la creación de la declaración sobre los derechos del cliente, Willis fue el primer corredor de seguros en tener un compromiso con la transparencia, rechazar las cuotas y de negarse a tomar compensaciones adicionales, teniendo una posición firme de actuar en el mejor interés del cliente.

La cultura One Flag asegura que todos los miembros, independientemente de su ubicación, forman parte de un equipo integrado a ayudar a los clientes. Este es el espíritu que inspira el servicio Glocal que ofrece los recursos globales de Willis en el mercado local y mueve el modelo Client Advocacy, en el cual un profesional especializado trae los conocimientos del mercado y crea una estrategia innovadora completa.

En todo el mundo, los 20,000 asociados de Willis demuestran su compromiso y pasión por lo que hacen y todas las acciones, tienen el cliente en primer lugar.

Willis en el Perú cuenta con una reconocida especialidad en clientes corporativos y con operaciones a nivel nacional. El equipo de más de 100 colaboradores puestos a disposición de nuestros clientes se encuentra

interconectado a la red de Willis Group con el permanente respaldo de nuestros especialistas ubicados en las distintas oficinas a nivel mundial.

La filosofía de trabajo y el equipo de especialistas altamente calificados, le permite atender las necesidades específicas de cada cliente con soluciones personalizadas, optimizando la inversión y promoviendo resultados.

### **4.3. LÍNEAS DE NEGOCIO**

#### **4.3.1. Riesgos humanos**

Willis pone a disposición de los clientes todo el conocimiento adquirido en Riesgos Humanos, el cual contempla los seguros de:

- Asistencia Médica
- EPS
- SCTR Salud
- SCTR Pensiones
- Vida
- Accidentes Personales.

Con una expresiva cartera de empresas nacionales y multinacionales Willis Perú es reconocido en el mercado porque identifica la mejor estrategia, optimiza la inversión, y promueve soluciones de acuerdo con la necesidad de cada cliente, todo esto con la mayor transparencia en el proceso.

La estructura de Beneficios está formada por profesionales altamente especializados, con el objeto de proporcionar un asesoramiento

personalizado y completo al personal del departamento de Recursos Humanos de cada uno de nuestros clientes.

Willis actúa aliado al cliente, con el propósito de reducir los costos y brindar asesoramiento en el mantenimiento de la calidad de vida de los beneficiarios, estudiando y entendiendo las necesidades particulares de nuestro cliente.

El compromiso asumido con el cliente convierte a Willis en un proveedor de soluciones en la gestión de beneficios, atendiendo las necesidades específicas y prestando total asesoría por medio de equipos integrados que actúan simultáneamente, buscando constantemente la forma de dar un servicio más eficiente y coordinado.

#### **4.3.2. Riesgos corporativos**

Willis provee servicios de Administración de Riesgos y Seguros para una amplia gama de empresas de diversos sectores económicos.

Su atención se sustenta sobre la base de soluciones creativas e innovadoras para la financiación de los riesgos. Cuenta con profesionales altamente capacitados en diferentes disciplinas que proveen asesoría en seguros tales como:

- Patrimoniales
- Responsabilidad Civil
- Aviación

- Transportes
- Logística
- Vehículos
- Financieros
- Crimen
- Construcción
- Energía
- Administración de Riesgos

#### **4.3.3. Grupos de afinidad**

La industria de grupos de afinidades es una de las fronteras más recientes y promisorias del mercado asegurador. Los llamados “Programas de Affinity” forman parte de una categoría que comparte las mismas necesidades especiales que ya desde hace algunos años constituyen un mercado en constante expansión.

Willis Affinity se especializa en la creación, planeamiento, distribución, desarrollo y gestión de programas de seguros masivos. La relación de Willis Affinity con el mercado financiero de seguros y marketing, alineada a su experiencia en el mercado asegurador refuerza y aumenta las ventas y desempeño financiero de nuestros clientes, frente a cualquier escenario.

Con base en su experiencia de mercado, su empresa puede:

- Generar renta adicional
- Mejorar los márgenes de utilidad
- Crear ventajas competitivas

- Atraer nuevos clientes y mejorar su retención
- Adicionar nuevos recursos (atributos) en los productos

Las acciones están basadas en entregar soluciones, optimizar procesos y generar servicios agregados con técnicas específicas y tecnología adecuada para las empresas de diversos segmentos, con creatividad, innovación y flexibilidad, con miras a atender las necesidades con calidad y excelencia, buscando siempre satisfacer a los clientes.

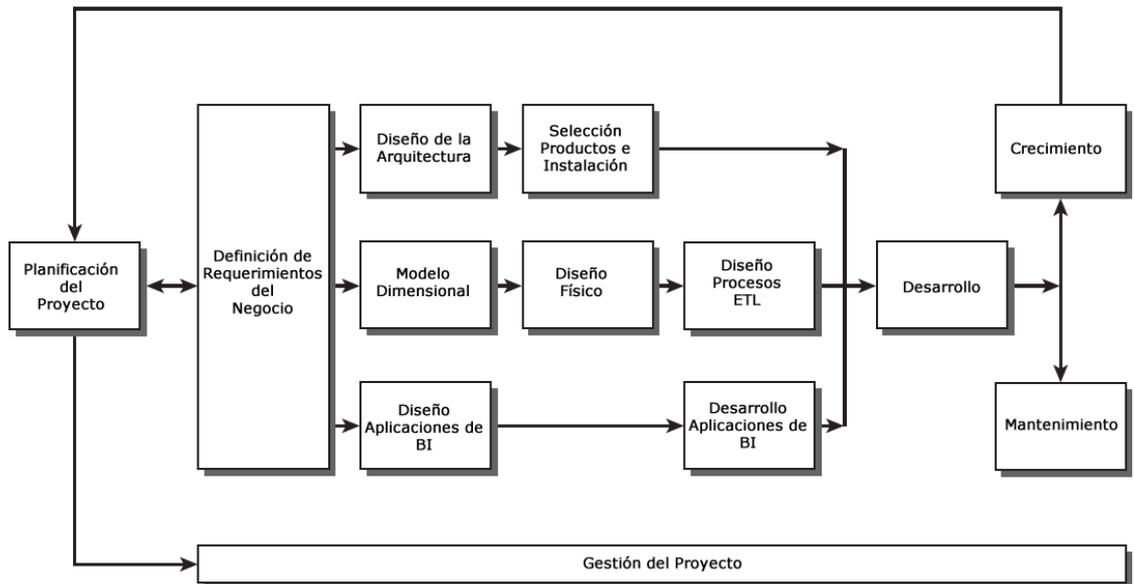
## **CAPÍTULO V**

### **IMPLEMENTACIÓN DEL MODELO**

Para el diseño del modelo de Inteligencia de Negocios propuesta, nos basamos del enfoque del ciclo de vida de Kimball, el cual se resume en la Figura N° 42. El diagrama muestra la secuencia de tareas, la dependencia y la concurrencia. Sirve como una hoja de ruta que nos ayudará a hacer lo correcto en el momento adecuado. El diagrama no refleja una línea de tiempo absoluta; Aunque los cuadros son igualmente anchos, hay una gran diferencia en el tiempo y el esfuerzo necesarios para cada actividad principal.

El enfoque del ciclo de vida de Kimball se enfoca en crear valor para los negocios, colaborar con el negocio además tener un desarrollo incremental. Sin embargo, también es fundamental que en el diseño y desarrollo de soluciones de Inteligencia de Negocios (BI) debe construirse sobre una base sólida de arquitectura de datos.

Figura N° 42: Ciclo de Vida del Proyecto



Fuente: The Data Warehouse Toolkit, 3<sup>rd</sup> Edition

## 5.1. PLANIFICACIÓN DEL PROYECTO

### 5.1.1. Objetivos del proyecto

Desarrollar un modelo de inteligencia de negocios basado en minería de opinión que permita realizar un análisis multidimensional de los comentarios vertidos en redes sociales que permita tomar decisiones a partir de información basado en hechos.

### 5.1.2. Definición del proyecto

El modelo de inteligencia de negocios a implementar permitirá brindar información relevante para la toma de decisiones a la gerencia de operaciones y comercial quienes han mostrado su interés por conocer el comportamiento que tienen los usuarios en redes sociales acerca de los servicios de corretaje de seguros que brinda la empresa.

El modelo de inteligencia de negocios a implementar permitirá extraer, transformar, clasificar e identificar sentimientos en los comentarios vertidos por los usuarios en redes sociales. Permitirá descubrir Insight en los comentarios como son: comportamientos y formas de expresión aplicando técnicas de la clasificación de texto e identificación de sentimientos aplicando técnicas de análisis de sentimientos.

El modelo de inteligencia de negocios será desarrollado utilizando las herramientas de inteligencia de negocios de Microsoft: construcción de paquetes de extracción, transformación y carga (ETL), motor de base datos relacional, base de datos dimensional, explotación de datos para la construcción de tableros de control, así como software, frameworks y librerías para entrenamiento y aprendizaje de máquina.

### **5.1.3. Alcance del proyecto**

Desarrollar un modelo inteligencia de negocios que permita realizar minería de opinión sobre redes sociales utilizando herramientas de la suite Microsoft Business Intelligence y software de aprendizaje de máquina.

Para el proceso de extracción, transformación y carga se utilizará las herramientas de SQL Server Integration Services la cual permite conectarnos a diferentes motores de base datos relacionales como: SQL Server Database Engine, archivos planos, hojas de cálculo, etc.

Para la construcción del almacén de datos se utilizará el servidor: SQL Server Análisis Services.

Para la construcción de los tableros de control y explotación de los datos se utilizará Microsoft Excel y Power BI.

## **5.2. REQUERIMIENTOS DEL NEGOCIO**

Los requerimientos que tiene el negocio son el de reducir el tiempo de atención que tienen los reclamos, conocer los temas o tópicos de los que se habla e identificar los sentimientos que tienen los usuarios en redes sociales acerca de los productos y servicios de la empresa corredora de seguros.

Los requerimientos, los antecedentes y el marco teórico están relacionadas directamente con los objetivos de la presente investigación.

## **5.3. MODELAMIENTO DIMENSIONAL**

El modelamiento dimensional es parte del ciclo de vida en el desarrollo de soluciones de Inteligencia de Negocios (BI) desarrollada por Ralph Kimball el cual incluye un conjunto de métodos, técnicas y conceptos estándares que se usaran en el diseño de la arquitectura de un data warehouse, el cual permitirá una alta performance para responder a las consultas de los usuarios finales a través de las aplicaciones de BI así como la actualización de los datos por los procesos de Extracción, Transformación y Carga (ETL) .

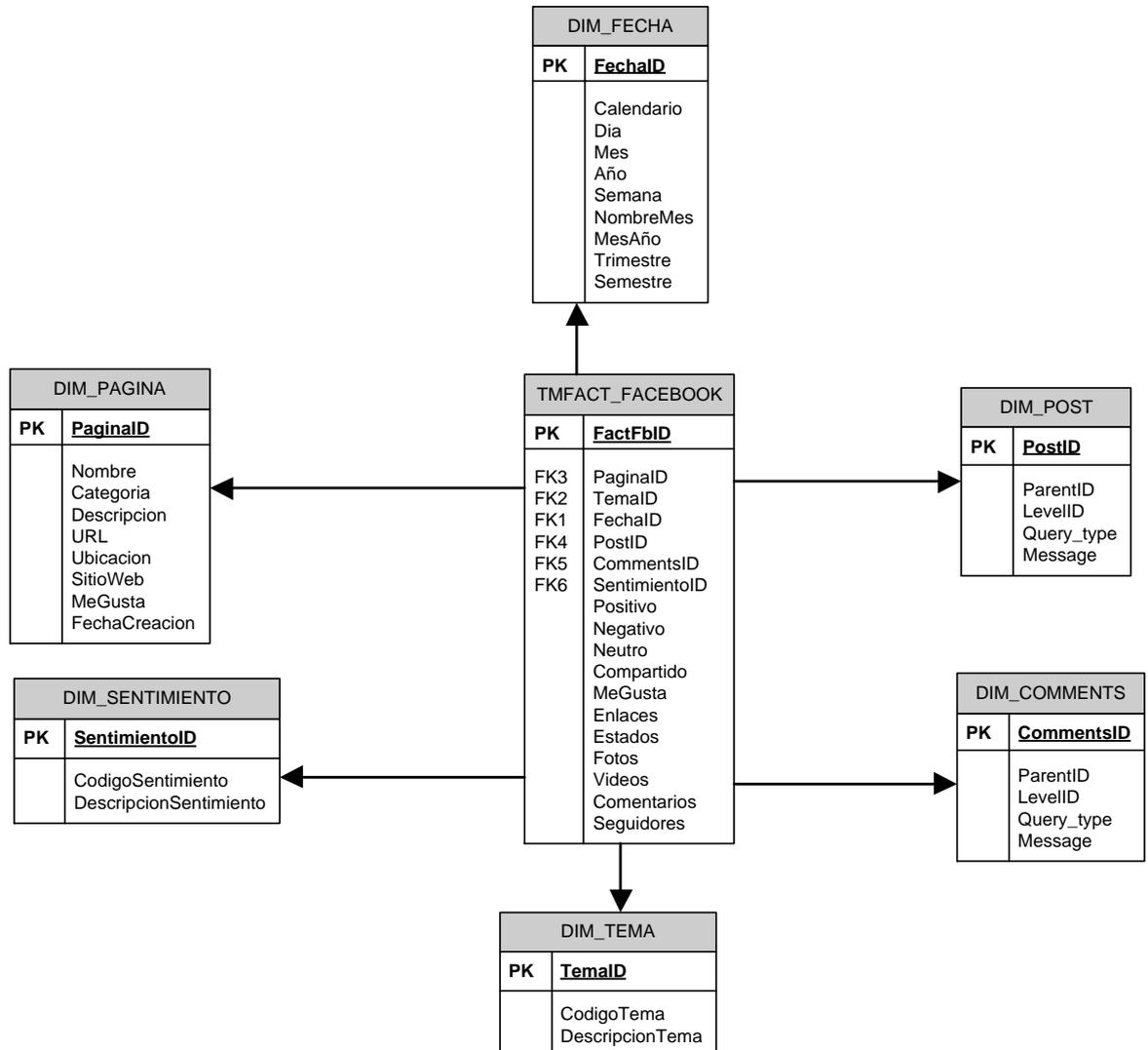
### **5.3.1. Modelo dimensional**

Luego de realizar el análisis dimensional de los datos que son extraídos de la plataforma Facebook, se ha diseñado el modelo dimensional. Como se puede ver en la Figura N° 43, el cual tendrá un esquema estrella

El modelo dimensional es una base de datos relacional con esquema estrella, la cual se compone de una tabla de hechos que se ubica en la parte central del modelo y a los lados se encuentran ubicadas las dimensiones como satélites, las cuales encuentran relacionados a la tabla de hechos a través de una clave subrogada.

Al hacer las combinaciones entre la tabla de hechos con cada una de las tablas de dimensiones, el motor de base de datos dimensional generará todas las combinaciones posibles que nos permitirá explorar en la información almacenada desde diferentes ángulos, permitiendo a los usuarios de negocio tomar mejores decisiones basadas en hechos.

Figura N° 43: Modelo Dimensional



Fuente: Elaboración Propia (2020)

### 5.3.1.1. Componentes del modelo dimensional

El modelo dimensional de Facebook está compuesto de los elementos básicos de un modelo dimensional con un esquema estrella. Como se puede apreciar en la Tabla N° 4, se encuentra detallada la estructura de la tabla de hechos y en las Tablas N° 5, 6, 7, 8, 9 y 10 la estructura de las tablas de dimensiones del modelo.

Tabla N° 4: Tabla de hechos: FACT\_FACEBOOK

Nombre Campo	Tipo Dato	Long.	Descripción Campo	Clave Primaria	Clave Foránea
FactFbID	Integer	8	Clave principal tabla de hechos	Si	No
PáginaID	Integer	4	Clave foránea tabla dimensión Página	No	Si
TemaID	Integer	4	Clave foránea tabla dimensión Tema	No	Si
FechaID	Integer	4	Clave foránea tabla dimensión Fecha	No	Si
SentimientoID	Integer	4	Clave foránea tabla Dimensión Sentimiento	No	Si
Positivo	Boolean	4	Reacción positiva	No	No
Negativo	Boolean	4	Reacción negativa	No	No
Neutro	Boolean	4	Reacción neutra	No	No
Compartido	Integer	8	Cantidad contenido compartido	No	No
MeGusta	Integer	8	Cantidad me gusta	No	No
Enlaces	Integer	8	Cantidad enlaces compartidos	No	No
Estados	Integer	8	Veces cambió estado	No	No
Fotos	Integer	8	Cantidad fotos subidas	No	No
Videos	Integer	8	Cantidad videos subidos	No	No
Comentarios	Integer	8	Cantidad comentarios	No	No
Seguidores	Integer	8	Cantidad seguidores	No	No

Fuente: Elaboración Propia (2019)

Tabla N° 5: Dimensión DIM\_TEMA

Nombre Campo	Tipo Dato	Long.	Descripción Campo	Clave Primaria	Clave Foránea
TemaID	Short	4	Clave sutituta de dimensión de la tabla	Si	No
Código	Varchar	10	Código de Tema	No	No
Descripción	Varchar	100	Descripción extendida del tema	No	No

Fuente: Elaboración Propia (2019)

Tabla N° 6: Dimensión DIM\_SENTIMIENTO

Nombre Campo	Tipo Dato	Long.	Descripción Campo	Clave Primaria	Clave Foránea
SentimientoID	Short	4	Clave sutituta de dimensión de la tabla	Si	No
Código	Varchar	10	Código de Sentimiento	No	No
Descripción	Varchar	100	Descripción extendida del Sentimiento	No	No

Fuente: Elaboración Propia (2019)

Tabla N° 7: Dimensión DIM\_FECHA

Nombre Campo	Tipo Dato	Long.	Descripción Campo	Clave Primaria	Clave Foránea
FechaID	Integer	8	Clave principal de la tabla	Si	No
Calendario	DateTime	8	Fecha dd/mm/aaaa	No	No
Día	Varchar	2	Día de la fecha	No	No
Mes	Varchar	2	Mes de la fecha	No	No
Año	Varchar	4	Año de la fecha	No	No
DíaSemana	Integer	4	Número día la semana	No	No
DíaMes	Integer	4	Número día del mes	No	No
SemanaAño	Integer	4	Número semana del año	No	No
MesAño	Integer	4	Número mes del año	No	No
Trimestre	Integer	4	Trimestre año	No	No
Semestre	Integer	4	Semestre año	No	No

Fuente: Elaboración Propia (2019)

Tabla N° 8: Dimensión DIM\_PAGINA

Nombre Campo	Tipo Dato	Long.	Descripción Campo	Clave Primaria	Clave Foránea
PaginaID	Integer	8	Clave Principal de la tabla	Si	No
Nombre	Varchar	100	Nombre de la página	No	No
Categoría	Varchar	100	Categoría asignada página	No	No
Descripción	Varchar	255	Descripción extendida página	No	No
URL	Varchar	100	Enlace de la página	No	No
Ubicación	Varchar	100	Ubicación de la página	No	No
SitioWeb	Varchar	100	Nombre del Sitio	No	No
MeGusta	Integer	8	Cantidad me gusta	No	No
FechaCreación	Date	8	Fecha de creación de la página	No	No

Fuente: Elaboración Propia (2019)

Tabla N° 9: Dimensión DIM\_PUBLICACION

Nombre Campo	Tipo Dato	Long.	Descripción Campo	Clave Primaria	Clave Foránea
PublicaciónID	Integer	8	Clave Principal de la tabla	Si	No
ObjectID	Varchar	11	ID del objeto	No	No
ParentID	Integer	8	ID del Padre	No	No
ID	Integer	8	ID Publicación	No	No
Publicación	Varchar	8000	Detalle de la publicación	No	No

Fuente: Elaboración Propia (2019)

Tabla N° 10: Dimensión DIM\_COMENTARIO

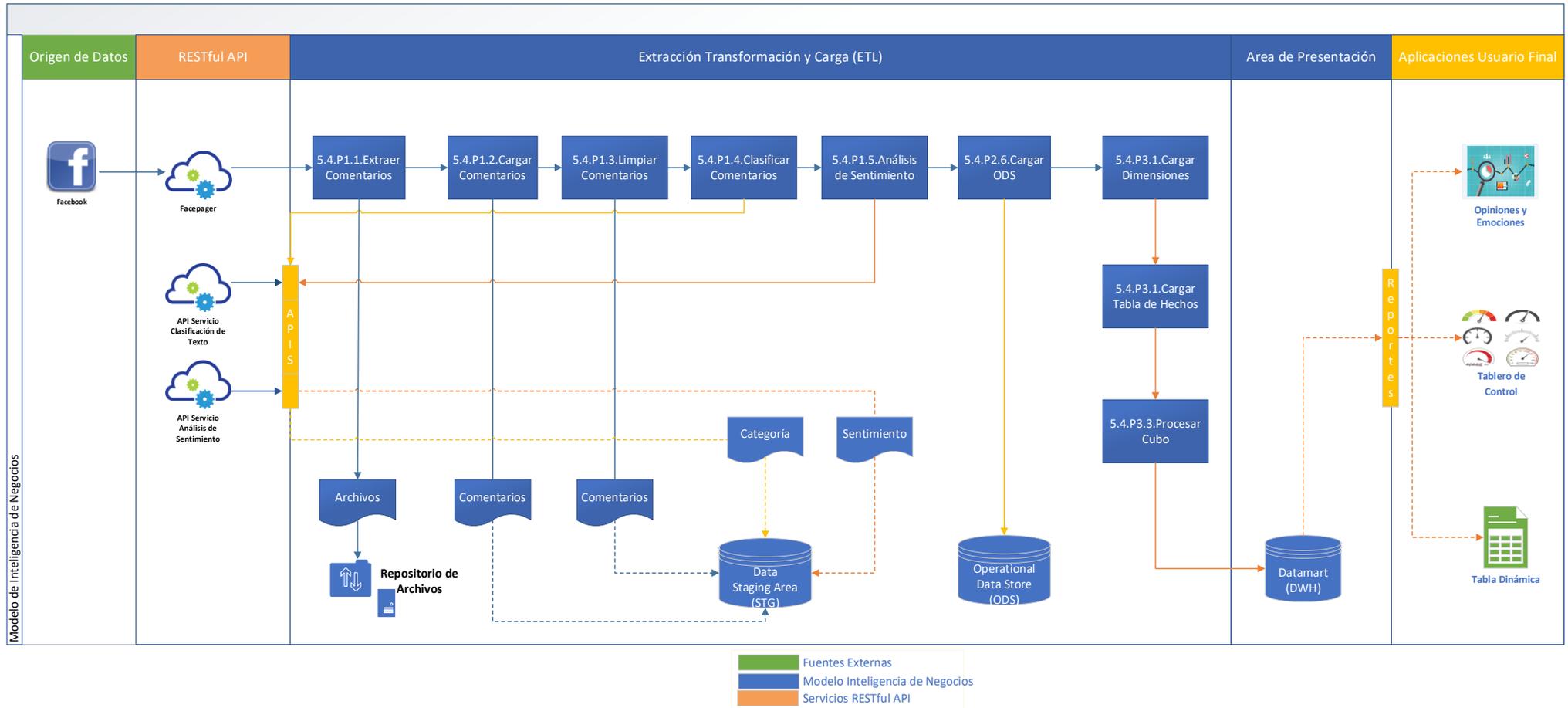
Nombre Campo	Tipo Dato	Long.	Descripción Campo	Clave Primaria	Clave Foránea
ComentarioID	Integer	8	Clave Principal de la tabla	Si	No
ObjectID	Varchar	11	ID del objeto	No	No
ParentID	Integer	8	ID del Padre	No	No
ID	Integer	8	ID Publicación	No	No
Comentario	Varchar	8000	Detalle del comentario	No	No

Fuente: Elaboración Propia (2019)

## 5.4. MODELO DE INTELIGENCIA DE NEGOCIOS

A través del siguiente modelo de Inteligencia de Negocios propuesta en la Figura N° 44, se puede visualizar a través de gráficos las diferentes etapas, procesos, componentes y relaciones que se debe completar para alcanzar los objetivos de la presente investigación.

Figura N° 44: Modelo de Inteligencia de Negocios

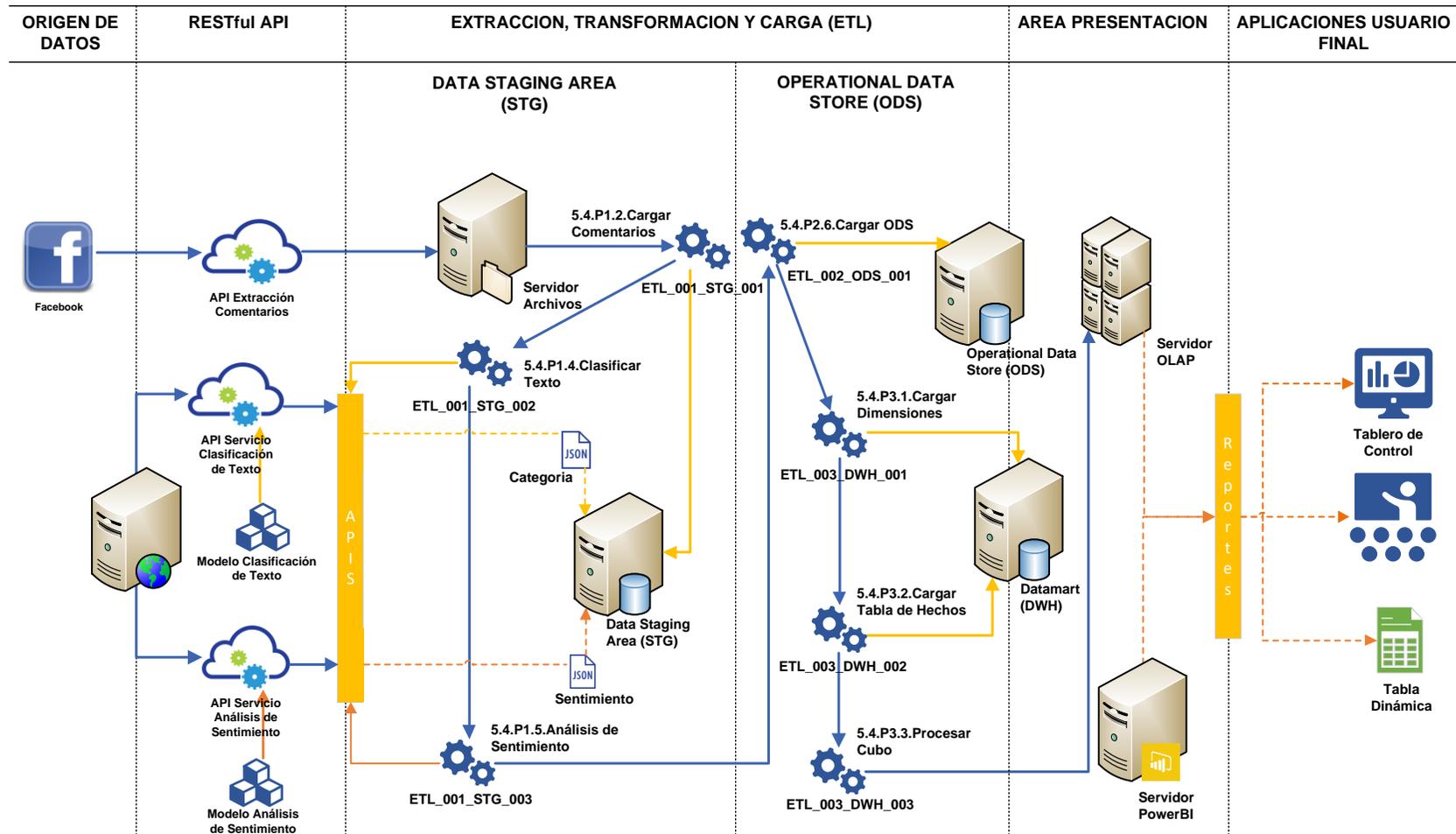


Fuente: Elaboración Propia (2021)

## 5.5. ARQUITECTURA DEL MODELO DE INTELIGENCIA DE NEGOCIOS

Después de realizar el análisis dimensional, el análisis de los datos y de los componentes de hardware y software necesarios del modelo, en la Figura N° 45 se puede visualizar la arquitectura que soportará la implementación del modelo de Inteligencia de Negocios.

Figura N° 45: Arquitectura del Modelo de Inteligencia de Negocios



Fuente: Elaboración Propia (2021)

## **5.6. HERRAMIENTAS DEL MODELO**

Para la construcción del modelo de Inteligencia de Negocios se utilizarán las herramientas de la suite Microsoft Business Inteligencia, las cuales son un conjunto de herramientas diseñadas especialmente para el desarrollo e implementación de soluciones de Inteligencia de Negocios (BI) y para el procesamiento de datos. Estas herramientas tienen como objetivo ayudar a las empresas a convertir toda la información compleja y dispersa en información extraíble.

Para la implementación del modelo de Inteligencia de Negocios propuesta, se ha previsto utilizar las siguientes herramientas que Microsoft dispone en la suite de Business Intelligence:

- SQL Server Database Engine
- SQL Server Data Tools (SSDT)
  - SQL Server Analysis Services (SSAS)
  - SQL Server Integration Services (SSIS)
- SQL Server Management Studio (SSMS)
- SQL Server Profiler

### **5.6.1. SQL Server Database Engine**

El motor de base de datos de SQL Server es el principal servicio para almacenar, procesar y proteger los datos. El motor de base de datos SQL Server proporciona acceso controlado y procesamiento rápido de transacciones para cumplir con los requerimientos que tienen las

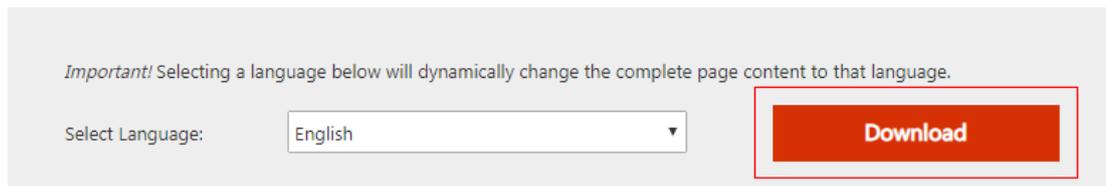
aplicaciones de operaciones de negocio, respecto al procesamiento, almacenamiento y consulta de datos.

Para poder instalar el motor de base datos SQL Server de manera correcta en sus servidores, se debe seguir las siguientes instrucciones:

#### **5.6.1.1. Descarga Microsoft SQL Server 2016**

Para descargar del archivo instalador del motor de base datos Microsoft SQL Server 2016, el usuario tiene que ir al siguiente enlace: <https://www.microsoft.com/en-us/download/details.aspx?id=56840>. Como se puede ver en la Figura N° 46, la nueva ventana muestra en la parte superior izquierda una lista desplegable que permite seleccionar el lenguaje del archivo instalador y en la parte superior derecha se muestra el botón [Download]. Para iniciar el proceso de descarga, el usuario deberá dar click sobre el botón [Download].

Figura N° 46: Descargar Microsoft SQL Server 2016



Microsoft® SQL Server® 2016 SP2 Express is a powerful and reliable free data management system that delivers a rich and reliable data store for lightweight Web Sites and desktop applications.

- [+ Details](#)

---

- [+ System Requirements](#)

---

- [+ Install Instructions](#)

---

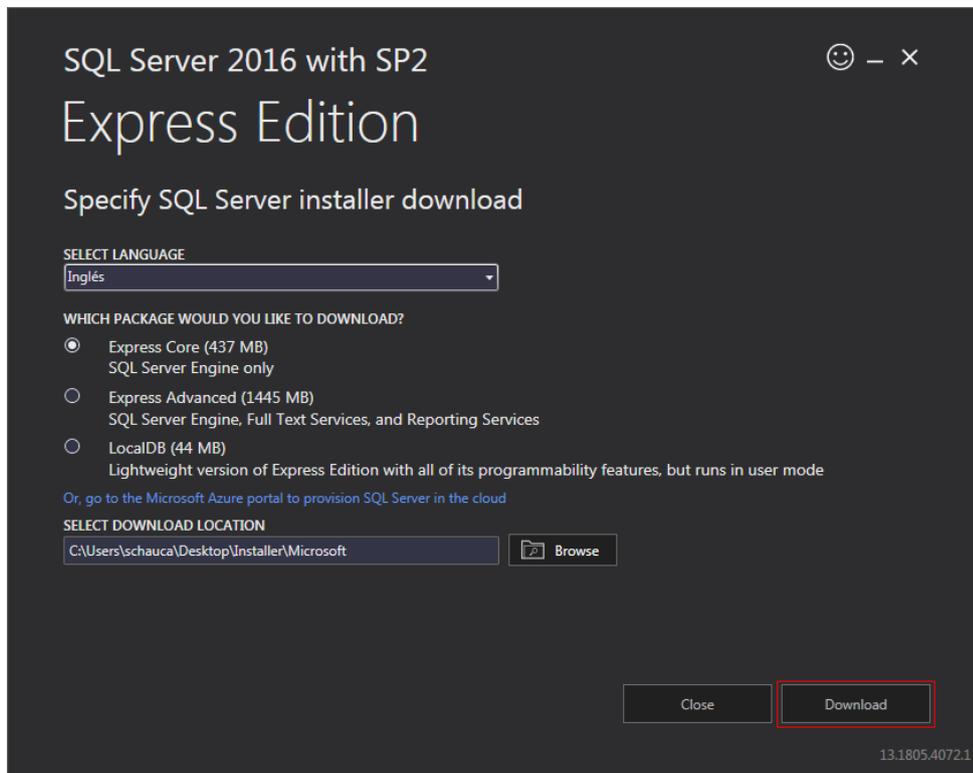
- [+ Additional Information](#)

---

#### 5.6.1.2. Instalar Microsoft SQL Server 2016

Para iniciar el proceso de instalación del motor de base de datos Microsoft SQL Server 2016 en nuestro equipo de desarrollo, primero se debe ubicar la ruta en la cual se encuentra ubicado el instalador descargado en el paso anterior, luego dar doble clic sobre el icono del instalador luego se abrirá una nueva ventana como se puede visualizar en la Figura N° 47, seleccionar los campos y completar la información requerida y finalmente clic sobre el botón [Download].

Figura N° 47: Instalar Microsoft SQL Server 2016



### 5.6.2. SQL Server Data Tools (SSDT)

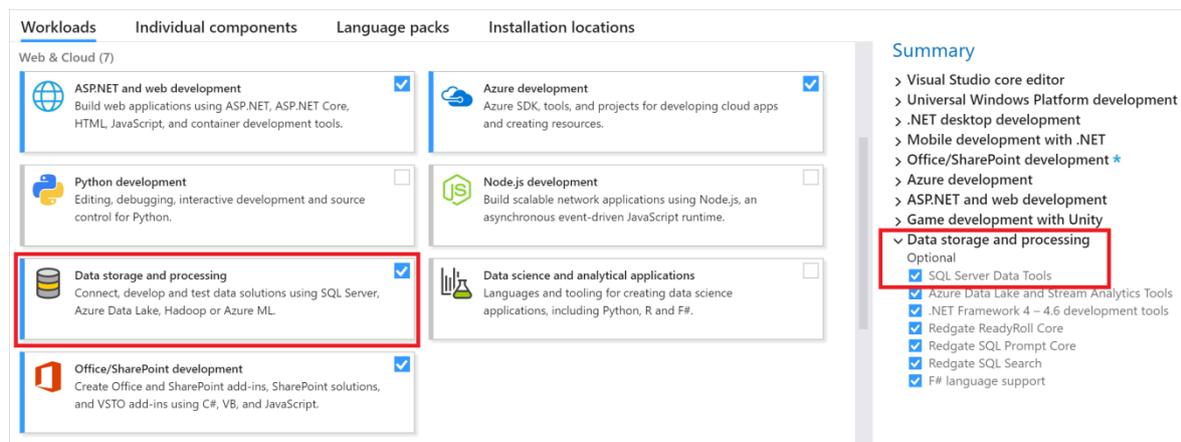
SQL Server Data Tools (SSDT) es una herramienta de desarrollo para construir bases de datos relacionales SQL, bases de datos SQL Azure, modelos de datos de Analysis Services (SSAS), paquetes de Integration Services (SSIS) y creación de reportes con Reporting Services (SSRS). Con SSDT, puede diseñar e implementar cualquier tipo de contenido de SQL Server con la misma facilidad con la que desarrollaría una aplicación en Visual Studio.

#### 5.6.2.1. Instalar SQL Server Data Tools (SSDT)

Como se puede apreciar en la Figura N° 48, para instalar la herramienta SQL Server Data Tools (SSDT) durante la instalación de Visual Studio, se

debe seleccionar la carga de trabajo (Work Load) de almacenamiento y procesamiento de datos, y luego seleccione Herramientas de datos de SQL Server. Si Visual Studio ya se encuentra instalado, puede editar la lista de cargas de trabajo para incluir SSDT:

Figura N° 48: Instalar SSDT con Visual Studio



Si desea descargar la versión standalone del instalador de SQL Server Data Tools (SSDT) directamente de la página de descargas de Microsoft, debe ir al siguiente link: <https://docs.microsoft.com/en-us/sql/ssdt/download-sql-server-data-tools-ssdt?view=sql-server-ver15#ssdt-for-vs-2017-standalone-installer>, luego de descargado el archivo instalador, se debe ejecutar el instalador dando doble click y siguiendo las instrucciones.

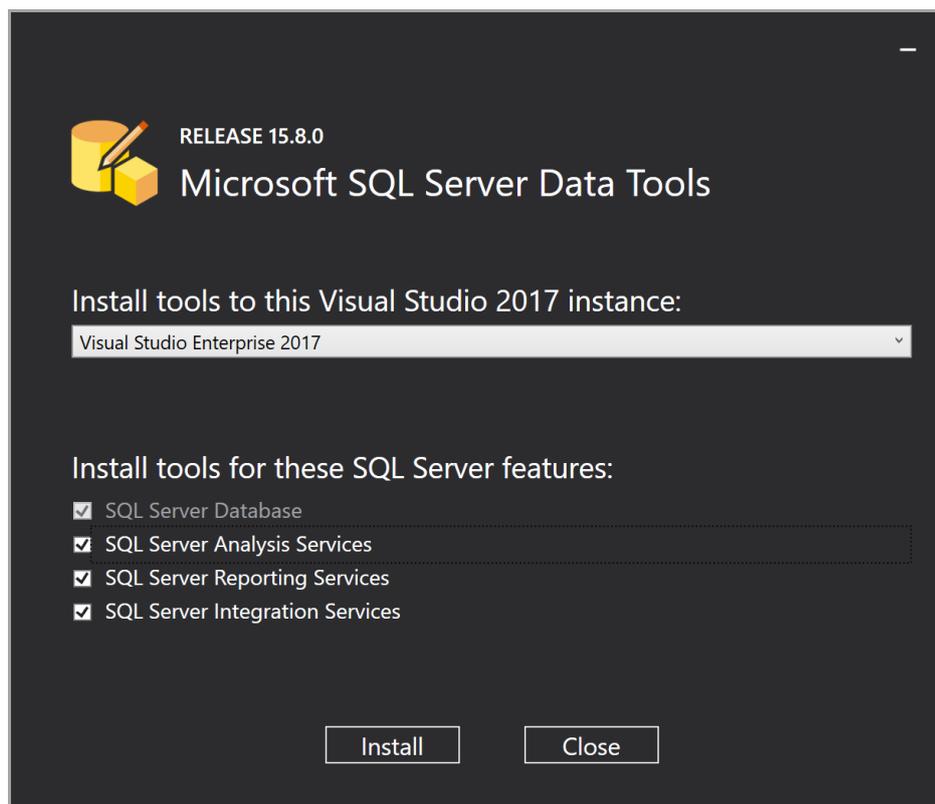
#### 5.6.2.2. SQL Server Analysis Services (SSAS), SQL Server Integration Services (SSIS) y SQL Server Reporting Services (SSRS)

Como se puede apreciar en la Figura N° 49, para instalar las herramientas de Inteligencia de Negocios desde instalador SSDT se deben habilitar las características disponibles: SQL Server Analysis Services (SSAS), SQL

Server Reporting Services (SSRS) y SQL Server Integration Services (SSIS), luego se debe dar click sobre el botón [Instalar].

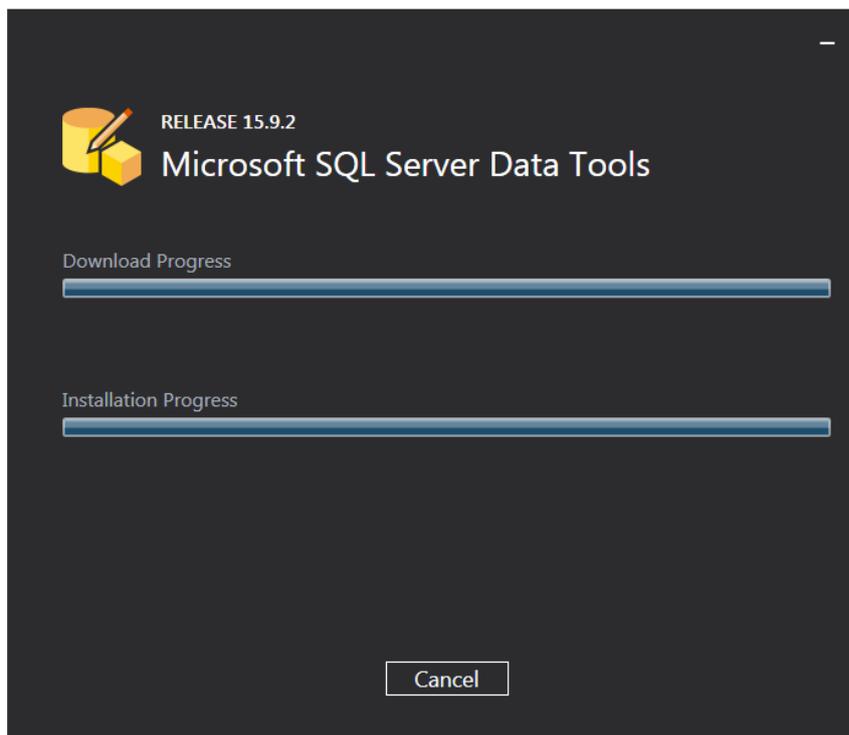
El instalador lista las instancias de Visual Studio disponibles para agregar las herramientas SSDT. Si Visual Studio aún no está instalado en el equipo de desarrollo, al seleccionar Instalar una nueva instancia de SQL Server Data Tools, se instala SSDT con una versión mínima de Visual Studio, pero para la mejor experiencia, se recomienda usar SSDT con la última versión de Visual Studio.

Figura N° 49: Instalar SSDT con soporte para SSAS, SSIS, SSRS



Como se puede apreciar en la Figura N° 50, se iniciará el proceso de instalación de cada una de las herramientas de soporte de proyectos de Inteligencia de Negocios (BI).

Figura N° 50: Instalación de SSDT



SQL Server Analysis Services (SSAS) es uno de los productos de la suite Microsoft Business Intelligence para desarrollar soluciones de procesamiento analítico en línea (OLAP). Es un motor de datos analíticos que es utilizado para la toma de decisiones y desarrollo de soluciones de Inteligencia de Negocios, proporciona los datos analíticos para reportes empresariales, aplicaciones cliente como Excel, Reporting Services y otras herramientas de Inteligencia de Negocios construida por terceros.

### 5.6.3. Facepager

#### 5.6.3.1. Descarga del Instalador Facepager

Los paquetes de instalación actualizados de la herramienta de extracción Facepager, se encuentran disponibles en la página de descarga que la herramienta Facepager tiene en la siguiente dirección: <https://github.com/strohne/Facepager/releases>. Como se puede apreciar en la Figura N° 51, en la página de descarga existen instaladores para los sistemas operativos: Windows y Mac OS.

Para iniciar la descarga se debe ubicar el enlace correspondiente al sistema operativo que se desee descargar y luego hacer click sobre este enlace.

Figura N° 51: Descarga del Instalador Facepager



Assets 4	
<a href="#">Facepager_4_3_10.app.macosx.zip</a>	230 MB
<a href="#">Facepager_Setup_4_3_10.windows.exe</a>	127 MB
<a href="#">Source code (zip)</a>	
<a href="#">Source code (tar.gz)</a>	

#### 5.6.3.2. Instalar Facepager en equipo cliente

Para iniciar el proceso de instalación de la herramienta de extracción de datos públicos desde las redes sociales, se debe ubicar la ruta en la cual se encuentra grabado el instalador que fue descargado en el paso anterior, luego se debe hacer doble click sobre el icono del instalador, se abrirá la ventana del asistente de instalación, luego se debe seleccionar las opciones requeridas y completar la información requerida.

## **5.7. CONSTRUCCIÓN DEL MODELO**

Para la construcción del modelo de Inteligencia de Negocios propuesta en la presente investigación, se deben realizar una serie de tareas previas las cuales son necesarias para definir la arquitectura de la solución. Como son la descarga e instalación de herramientas del modelo, el desarrollo de los procesos de: Extracción de datos desde Facebook, Clasificación de Texto, Análisis de Sentimiento, diseño del almacén de datos, construcción del proyecto de Integration Services, construcción del proyecto de Analysis Services, construcción del tablero de control en Power BI.

### **5.7.1. Extracción, transformación y carga (ETL)**

Facebook tiene una gran cantidad de información sobre personas, empresas, eventos, álbum de fotos, También tiene información de la forma cómo están relacionadas entre sí: qué persona posee cada álbum, qué personas aparecen en cada foto, qué empresa organiza cada evento, etc.

Obtener esta gran cantidad de información es una de las etapas fundamentales del proceso de extracción de la solución que estamos planteando. Para lo cual tenemos que entender el funcionamiento de la Interfaz de Programación de Aplicaciones (API Graph) y de la herramienta de extracción Facepager.

La recolección de los datos se realizará directamente desde la plataforma de las redes sociales en Facebook a través de las API Graph que se encuentran disponibles en la plataforma de redes sociales.

La recuperación de los datos se realizará periódicamente a través de programas rastreadores que serán construidos con lenguaje de programación Python, el cual permitirá consumir fácilmente las API publicadas por las plataformas de redes sociales. Finalmente, los datos serán almacenados en una base datos de redes sociales como datos en bruto y sin procesar.

#### **5.7.1.1. Extracción de comentarios usando API Graph**

La API Graph es la herramienta principal proveída por Facebook para que las aplicaciones construidas por los desarrolladores puedan realizar tareas de lectura y escritura en la gráfica social de Facebook. Todos los productos y SDK interactúan con la API Graph, así como el resto de las API son extensiones de la API Graph.

Entender el funcionamiento de la API Graph es fundamental para el desarrollo del modelo de extracción, por lo cual explicaremos en forma detallada cada uno de los pasos necesarios según la documentación de la API Graph la cual se encuentra disponible en la plataforma de desarrollador de Facebook.

##### **5.7.1.1.1. Registrar cuenta de usuario desarrollador**

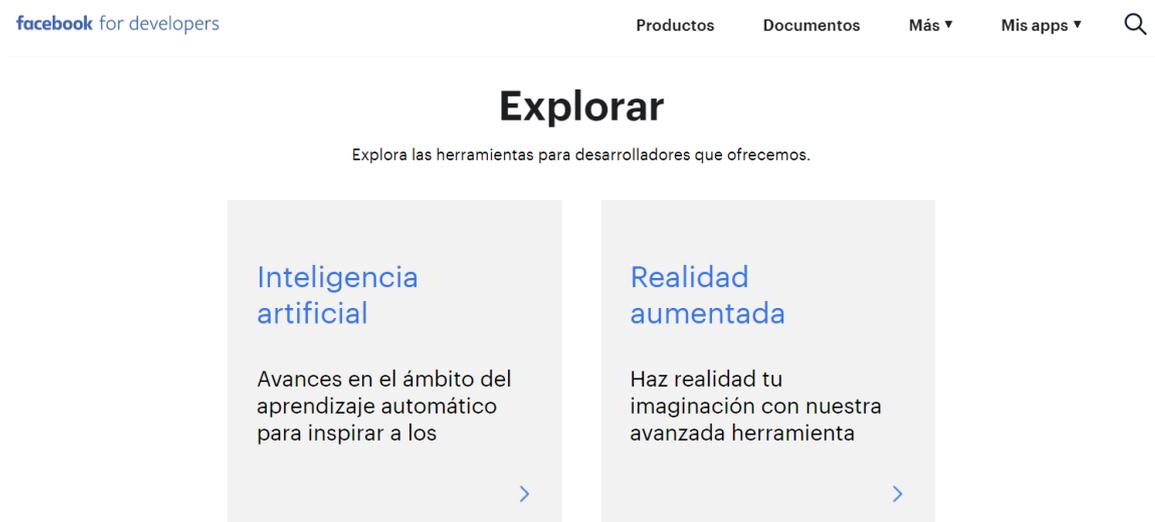
Para iniciar el proceso de extracción de los datos publicados por los usuarios en la plataforma Facebook, es necesario iniciar sesión con una cuenta de usuario. Cómo se puede ver en la Figura N° 52 y 53, lo primero que debe hacer en la página Facebook para desarrolladores:

<https://developers.facebook.com/>. es registrarse como un usuario desarrollador.

Figura N° 52: Registrar Cuenta de Usuario Desarrollador



Figura N° 53: Facebook para Desarrolladores



#### 5.7.1.1.2. Explorador de la API Graph

Para explorar la API Graph se debe ir al link de Facebook para desarrolladores: <https://developers.facebook.com/>. Luego desde el menú

principal, seleccionar la opción: [Herramientas], al seleccionar esta opción se desplegará un submenú con las opciones disponibles, finalmente como se muestra en la Figura N° 54, se debe seleccionar la opción: [Explorador de la API Graph].

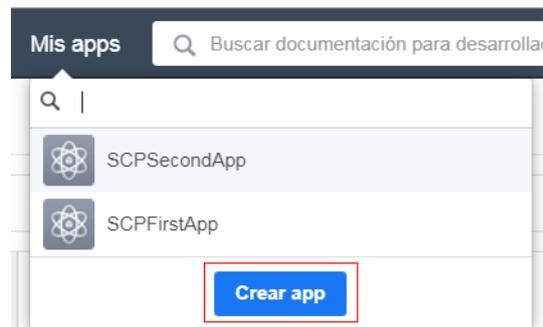
Figura N° 54: Explorador de la API Graph



### 5.7.1.1.3. Crear nueva aplicación (APP)

Para crear una nueva aplicación (APP) se debe ir al link del Explorador de la API Graph: <https://developers.facebook.com/tools/explorer>. Luego desde el menú principal seleccionar la opción: [Mis apps]. Como se puede apreciar en la Figura N° 55, al seleccionar esta opción se desplegará una ventana emergente con un listado de aplicaciones que el usuario desarrollador ya ha creado con su perfil y en la parte inferior de la ventana se muestra el botón [Crear app].

Figura N° 55: Crear Nueva Aplicación (APP)



Para iniciar el proceso de crear una nueva APP, se debe dar click sobre el botón [Crear app], Como se muestra en la Figura N° 56, luego se abrirá una ventana emergente con los campos de texto requeridos como son: [Nombre de la aplicación], [Correo electrónico de contacto] y en la parte inferior de la ventana se muestran los botones [Cancelar] y [Crear identificador de la App].

Para crear un nuevo identificador de aplicación (APP) primero se deben llenar los campos texto y dar click sobre el botón: [Crear identificador de la App]

Figura N° 56: Crear Nuevo Identificador de Aplicación (APP)

A screenshot of a form titled 'Crear un nuevo identificador de la app'. Below the title, there is a subtitle: 'Empieza a integrar Facebook en tu app o sitio web'. The form contains two text input fields: the first is labeled 'Nombre para mostrar' and has a placeholder text 'El nombre con el que quieres asociar este identificador de la app'; the second is labeled 'Correo electrónico de contacto'. At the bottom of the form, there is a horizontal line, followed by the text 'Al continuar, aceptas las Políticas de la plataforma de Facebook'. To the right of this text are two buttons: a grey 'Cancelar' button and a blue 'Crear identificador de la app' button, which is highlighted with a red rectangular border.

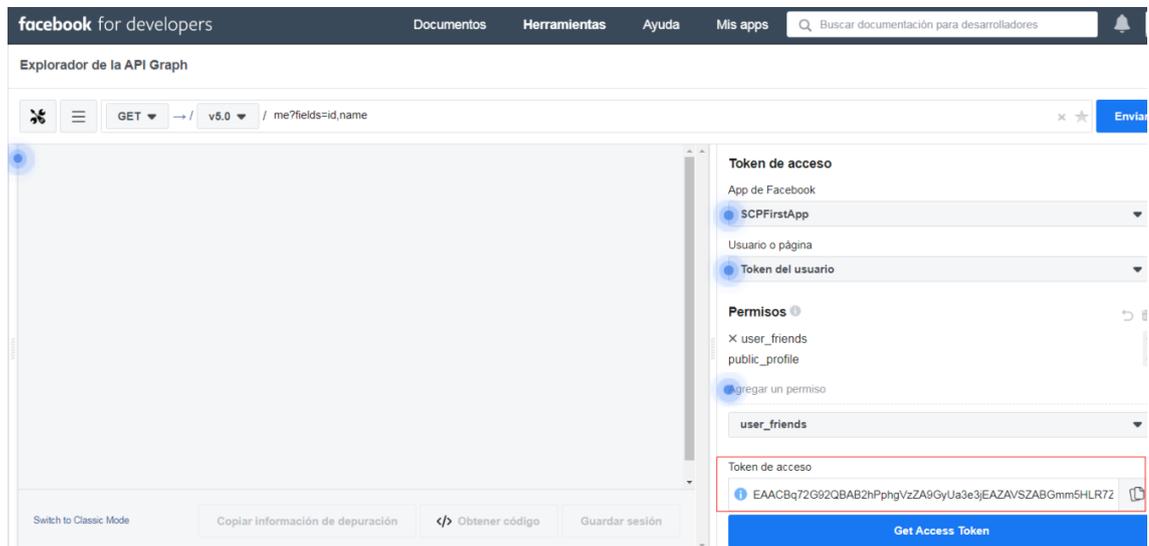
#### **5.7.1.1.4. Obtener token de acceso**

Un token de acceso es una cadena que identifica a un usuario, una app o una página, y que la app se puede utilizar para realizar llamadas a la API Graph. Cuando alguien se conecta con una app mediante el inicio de sesión con Facebook y aprueba la solicitud de permisos, la app obtiene un token de acceso que proporciona acceso temporal y seguro a las API de Facebook. Los tokens de acceso se obtienen con distintos métodos.

El token incluye información acerca de su caducidad y de la app que lo generó. Para efectos de las comprobaciones de privacidad, la mayoría de las llamadas a la API en Facebook deben incluir un token de acceso.

El token de acceso es uno de los parámetros necesarios para ejecutar una consulta a través de la API Graph. Como se puede ver en la Figura N° 57, para obtener el Token de Acceso se debe dar click sobre el botón: [Get Access Token]

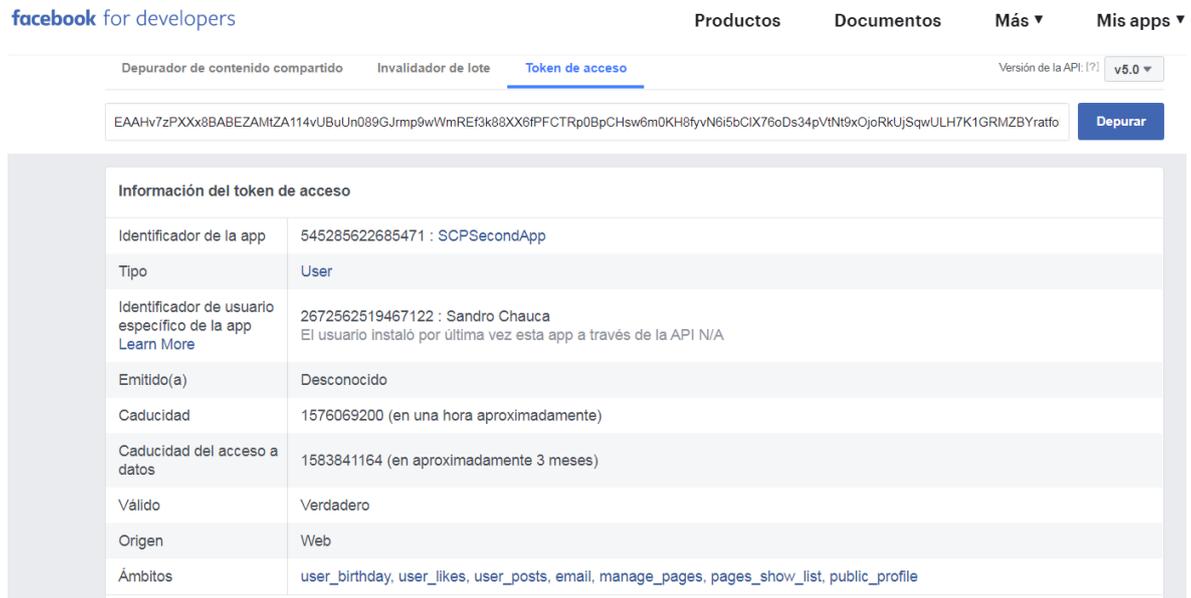
Figura N° 57: Obtener Token de Acceso



#### 5.7.1.1.5. Extender caducidad token de acceso

Como se puede apreciar en la Figura N° 58, el Token de Acceso tiene un tiempo de vida muy corto, por esta razón se debe solicitar la extensión del tiempo de caducidad. Para solicitar la extensión del la caducidad del Token de Acceso, primero se debe ir al siguiente link: <https://developers.facebook.com/tools/accesstoken/>

Figura N° 58: Extender Caducidad del Token de Acceso



The screenshot shows the 'Token de acceso' page in the Facebook for developers interface. At the top, there are navigation links for 'facebook for developers', 'Productos', 'Documentos', 'Más', and 'Mis apps'. Below these, there are tabs for 'Depurador de contenido compartido', 'Invalidador de lote', and 'Token de acceso'. A search bar contains a long alphanumeric string, and a 'Depurar' button is visible. The main content is a table titled 'Información del token de acceso' with the following data:

Información del token de acceso	
Identificador de la app	545285622685471 : SCPSecondApp
Tipo	User
Identificador de usuario específico de la app <a href="#">Learn More</a>	2672562519467122 : Sandro Chauca El usuario instaló por última vez esta app a través de la API N/A
Emitido(a)	Desconocido
Caducidad	1576069200 (en una hora aproximadamente)
Caducidad del acceso a datos	1583841164 (en aproximadamente 3 meses)
Válido	Verdadero
Origen	Web
Ámbitos	user_birthday, user_likes, user_posts, email, manage_pages, pages_show_list, public_profile

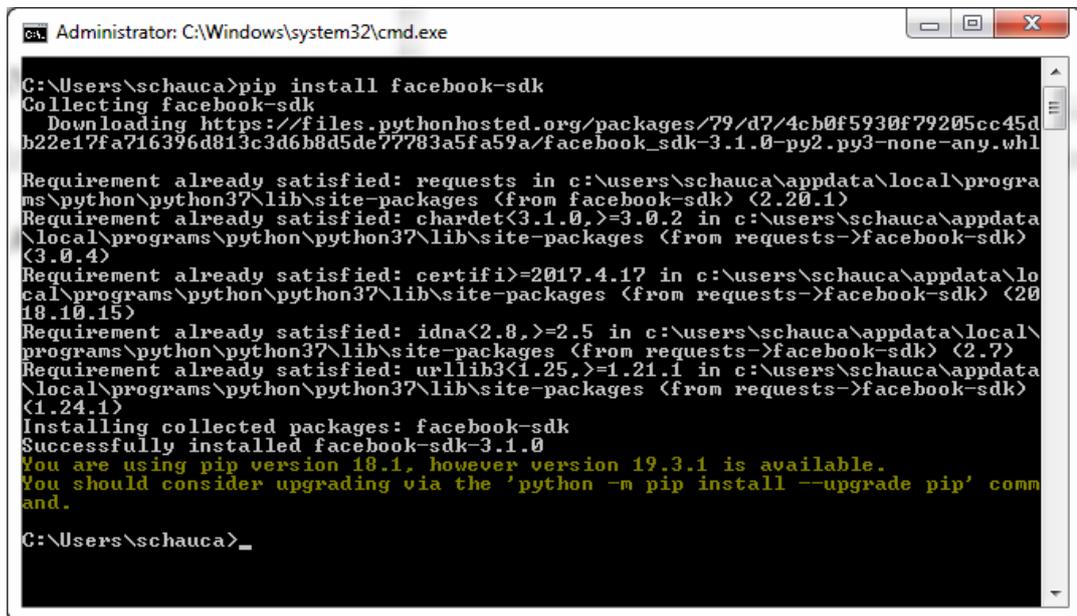
#### 5.7.1.1.6. Acceder a datos publicos con Python

Para poder acceder a los datos públicos de Facebook usando el lenguaje de programación Python, se debe instalar en el equipo del desarrollador el paquete: facebook-sdk. Para instalar el paquete facebook-sdk se debe abrir la ventana de comandos de windows y ejecutar el siguiente comando:

```
pip install facebook-sdk
```

Como se puede apreciar en la Figura N° 59 luego de ejecutar el commando de instalación pip, se iniciará el proceso de instalación de los paquetes necesarios del complemento Facebook-sdk, esperar hasta recibir el mensaje informativo que la instalación ha culminado satisfactoriamente.

Figura N° 59: Instalar Paquete Facebook-sdk



```
Administrator: C:\Windows\system32\cmd.exe
C:\Users\schauca>pip install facebook-sdk
Collecting facebook-sdk
  Downloading https://files.pythonhosted.org/packages/79/d7/4cb0f5930f79205cc45d
b22e17fa716396d813c3d6b8d5de77783a5fa59a/facebook_sdk-3.1.0-py2.py3-none-any.whl
Requirement already satisfied: requests in c:\users\schauca\appdata\local\progra
ms\python\python37\lib\site-packages (from facebook-sdk) (2.20.1)
Requirement already satisfied: chardet<3.1.0,>=3.0.2 in c:\users\schauca\appdata
\local\programs\python\python37\lib\site-packages (from requests->facebook-sdk)
(3.0.4)
Requirement already satisfied: certifi>=2017.4.17 in c:\users\schauca\appdata\lo
cal\programs\python\python37\lib\site-packages (from requests->facebook-sdk) (20
18.10.15)
Requirement already satisfied: idna<2.8,>=2.5 in c:\users\schauca\appdata\local\
programs\python\python37\lib\site-packages (from requests->facebook-sdk) (2.7)
Requirement already satisfied: urllib3<1.25,>=1.21.1 in c:\users\schauca\appdata
\local\programs\python\python37\lib\site-packages (from requests->facebook-sdk)
(1.24.1)
Installing collected packages: facebook-sdk
Successfully installed facebook-sdk-3.1.0
You are using pip version 18.1, however version 19.3.1 is available.
You should consider upgrading via the 'python -m pip install --upgrade pip' comm
and.
C:\Users\schauca>_
```

Para extraer los datos públicos de un usuario Facebook usando el lenguaje de programación Python, primero se debe enviar una consulta a la API Graph. La API Graph, recibirá los parámetros de ingreso, los procesará y si no encuentra ningún error de acceso, finalmente retornará un objeto con notación Json.

Como se puede visualizar en la Figura N° 60, en un editor de texto, escribir las líneas del programa de extracción. Primero se deben importar las librerías de trabajo: json y facebook para que puedan ser usados en todo el programa, además se deben proveer los parámetros de ingreso requeridos por la API Graph, como son: token de acceso y lista de campos.

Luego se debe crear una instancia de la clase GraphAPI() pasando como parámetro de ingreso el token de acceso, la instancia de la clase se asignará al objeto local graphAPI. Luego mediante el objeto graphAPI se ejecutará el

método `get_object()`, pasando como parámetro de ingreso el nodo y la lista de campos a consultar.

Figura N° 60: Acceder a datos públicos de Facebook con Python

```
import json
import facebook

def main():
    tokenAcceso =
    {"EAAHv7zPXXx8BABEZAMtZA114vUBuUn089GJrmp9wWmREf3k88XX6fPFCTRp0BpCHsw6m0
    KH8fyvN6i5bClX76oDs34pVtNt9xOjORkUjSqwULH7K1GRMZBYratfol2malyDM5roUnSZCU
    C4fr9oOrARobCwwtQWbVZBU6S1ZCd5ZAUtCggSYCdw9FkkZBXsNEQPQkrbtuiVsBgZDZD"}

    graphAPI = facebook.GraphAPI(tokenAcceso)
    listaCampos = ['email', 'name', 'last_name']
    perfil = graphAPI.get_object('me', fields = listaCampos)
    print(json.dumps(perfil, indent=4))

if __name__ == '__main__':
    main()
```

Luego de ejecutar el método `get_object()`, recibimos la respuesta del método invocado en el objeto `perfil`, el cual almacenará los valores de los campos definidos en el array: `listaCampos` como propiedades o atributos.

Finalmente, haciendo uso de la librería especializada: `json`, se procesa el objeto: `perfil` y se imprimen los atributos de este objeto con una notación: `Json`, como se muestra en las Figuras N° 61, 62 y 63.

Figura N° 61: Respuesta datos básicos en notación `Json`

```
{
  "email": "developer.tesis@hotmail.com",
  "name": "Developer",
  "last_name": "Tesis",
  "id": "0123456789012345"
}
```

Figura N° 62: Respuesta Likes en notación Json

```
{
  "likes" : {
    "data": [
      {
        "name": "Microsoft Developer",
        "id": "147590908618094",
        "created_time": "2015-10-07"
      }
    ]
  }
}
```

Figura N° 63: Respuesta Posts en notación Json

```
{
  "posts": {
    "data": [
      {
        "message": "Presentación SAP Business One HANA, el ERP
más rápido del mundo en Hotel Business Tower",
        "created_time": "2013-08-16",
        "id": "2672562519467122_574566072600121"
      }
    ]
  }
}
```

### 5.7.1.2. Extracción de comentarios con Facepager

Para el desarrollo del proceso de extracción de los datos públicos que se encuentran disponibles en las plataformas de redes sociales como: Facebook, se ha utilizado la herramienta de recolección de datos denominada: Facepager.

#### 5.7.1.2.1. Diseño de ventana principal de Facepager

Como se puede apreciar en la Figura N° 64, la herramienta Facepager está compuesta de 5 ventanas y una barra de menús con las cuales el usuario puede configurar consultas, configurar columnas, visualizar los resultados de la consulta a través de los Nodos, visualizar las columnas y ejecutar comandos para recolectar datos de las redes sociales:

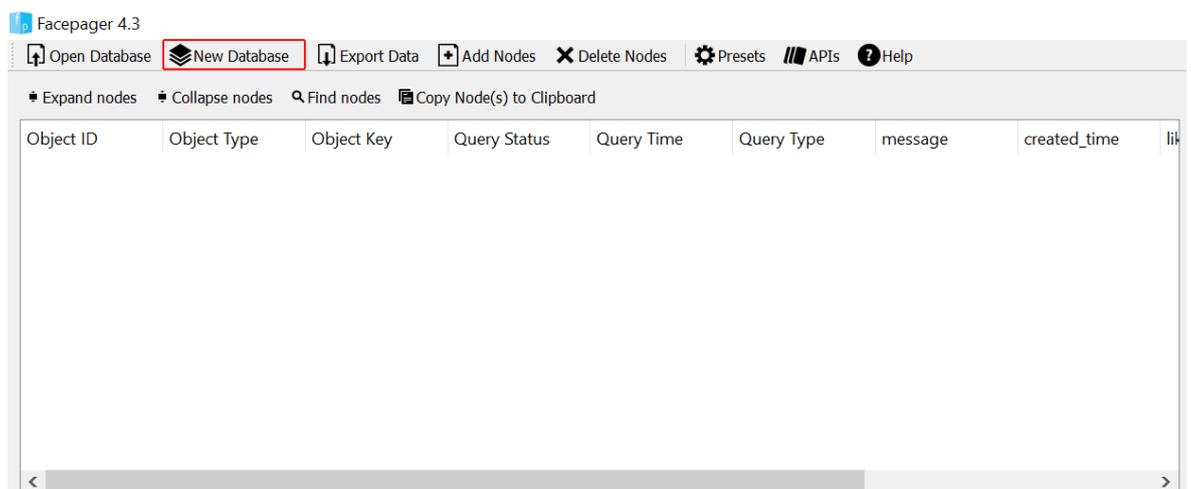
Figura N° 64: Diseño de la Ventana Principal en Facepager



### 5.7.1.2.2. Definición de la base datos en Facepager

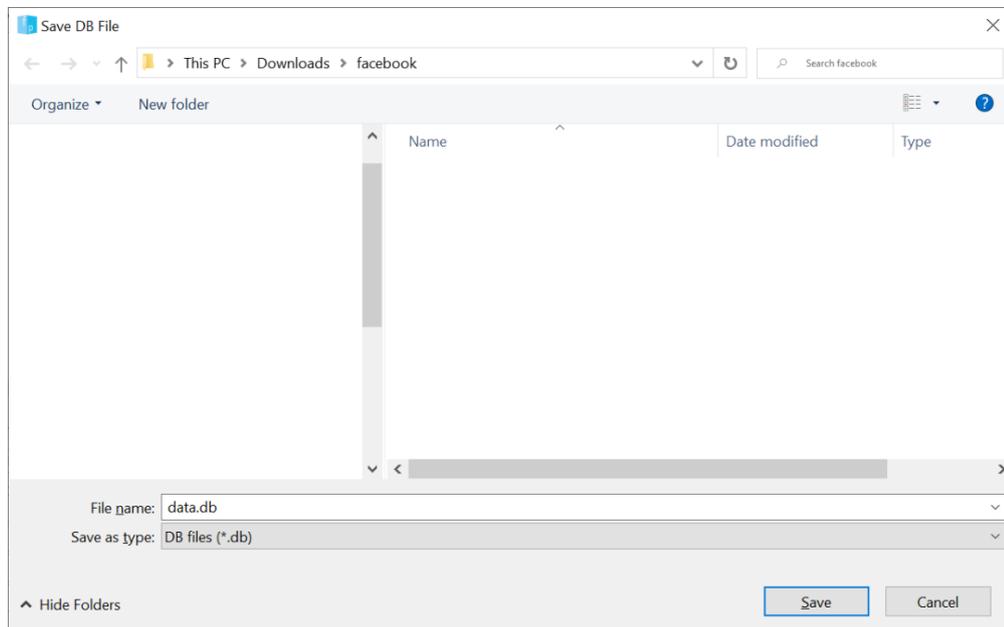
Para definir la base datos del proyecto en Facepager primero se debe ubicar la barra de menús en la ventana principal de la aplicación, como se puede apreciar en la Figura N° 65, luego se debe hacer click sobre el botón [New Database].

Figura N° 65: Base Datos del Proyecto en Facepager



Como se puede visualizar en la Figura N° 66, inmediatamente se desplegará la ventana que permitirá al usuario seleccionar la ruta en la cual se guardará el archivo de base datos, además de definir el nombre del archivo de base datos del proyecto: data.db

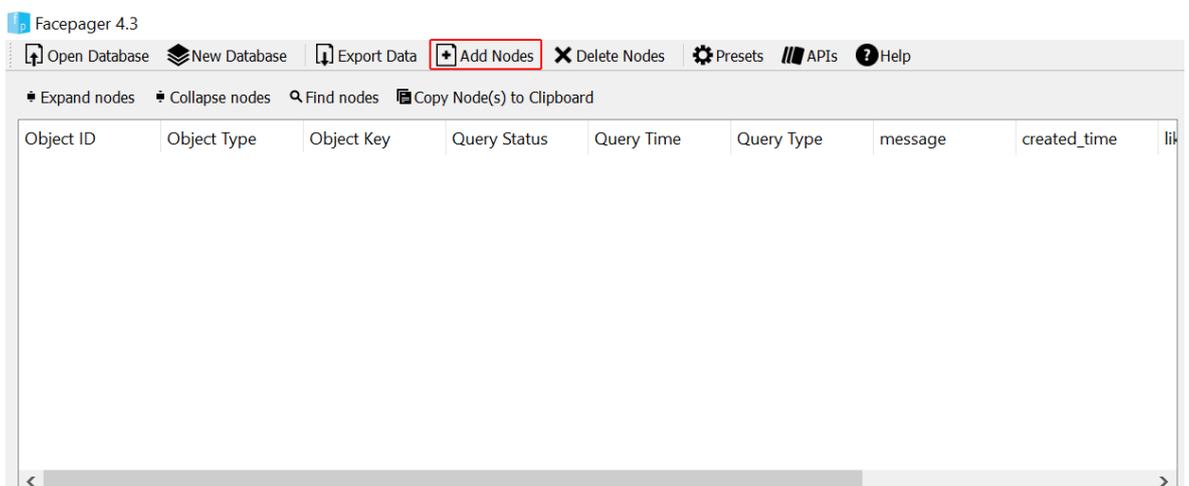
Figura N° 66: Ventana de Creación de Base Datos en Facepager



### 5.7.1.2.3. Definición de nodos de recolección de datos

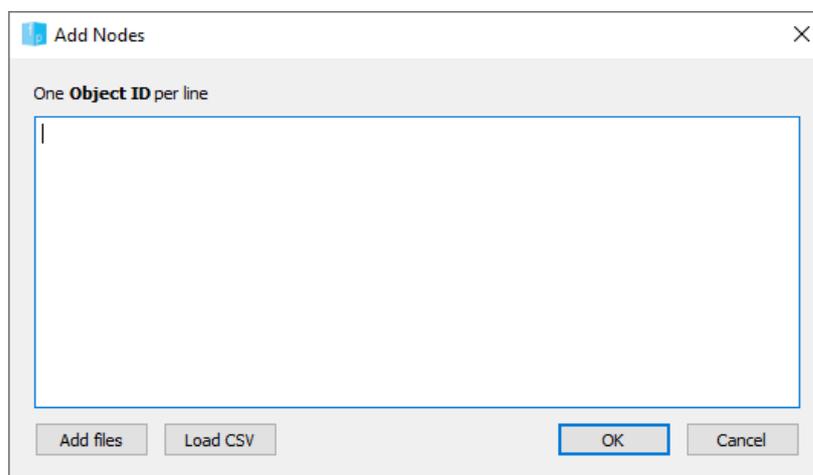
Para definir los nodos de recolección de datos del proyecto primero se debe ubicar la barra de herramientas de la aplicación, como se puede apreciar en la Figura N° 67 se encuentra en la parte superior de la ventana principal de la aplicación, luego hacer click sobre el botón: [Add Nodes]

Figura N° 67: Ventana de Definición de Nodos de Recolección de Datos



Como se puede apreciar en la Figura N° 68, inmediatamente se desplegará la ventana que permitirá añadir los nodos correspondientes a las cuentas de fanpages que serán extraídas a través del proceso de extracción.

Figura N° 68: Ventana de Definición de Nodos de Recolección de Datos



#### 5.7.1.2.4. Selección de módulos de extracción

Como se puede apreciar en la Figura N° 69, la herramienta Facepager dispone de una serie de módulos que permite la extracción de datos públicos desde las plataformas de redes sociales: YouTube, Twitter, Facebook, Amazon, Generic.

En la solución propuesta seleccionaremos el módulo de extracción de Facebook, el cual nos permite obtener los datos a través de la API Graph.

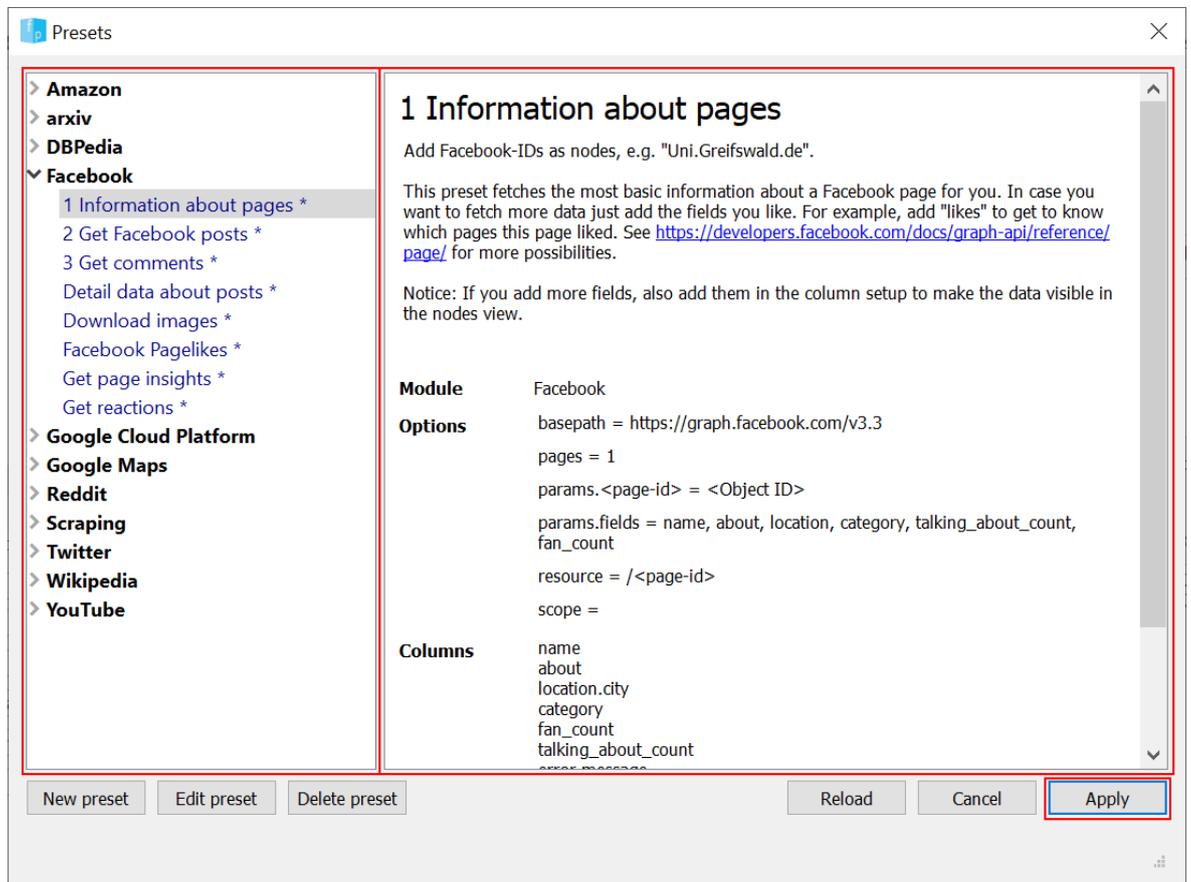
Figura N° 69: Módulos de Recolección de Datos

The image shows a web interface for configuring data collection modules. At the top, there are tabs for 'YouTube', 'Twitter', 'Twitter Streaming', 'Facebook' (which is highlighted with a red border), 'Amazon', and 'Generic'. Below the tabs, the 'Facebook' configuration section includes:

- 'Base path': A text input field.
- 'Resource': A dropdown menu with a '?' icon to its right.
- 'Parameters': A table with four rows. Each row has a dropdown menu on the left and a text input field on the right, with a '..' icon to the right of each input field.
- 'Maximum pages': A dropdown menu with the value '1' selected.
- 'Access token': A text input field.
- 'Settings': A button.
- 'Login to Facebook': A button.

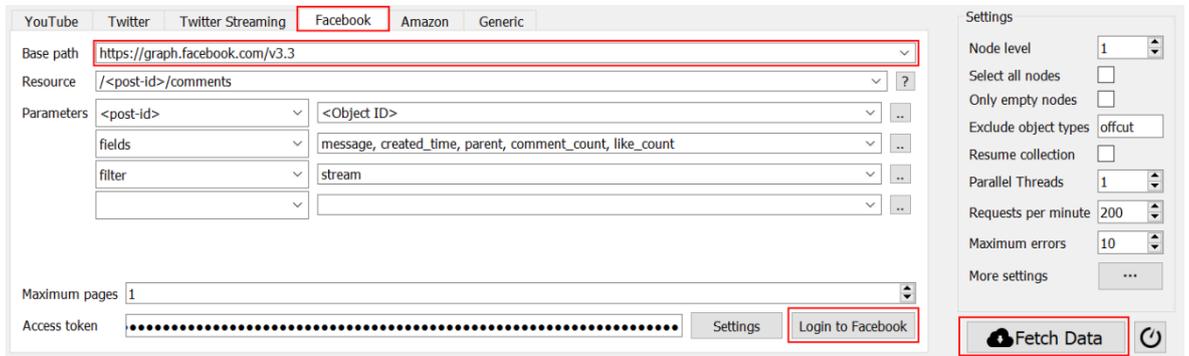
Para iniciar el proceso de extracción utilizaremos la ventana de Presets, como se puede visualizar en la Figura N° 70, nos muestra un panel izquierdo con módulos y los presets ya definidos en la herramienta, en el panel derecho la información de las opciones y las columnas del preset seleccionado y en la parte baja de la ventana el botón [Apply] el cual nos permitirá configurar la consulta con los valores del preset seleccionado.

Figura N° 70: Presets de Extracción de Datos



Para extraer los datos de comentarios de usuarios desde Facebook utilizaremos el Preset: “3 Get Comments”, como se puede apreciar en la Figura N° 71, los parámetros de la consulta han sido poblados por los valores del presets seleccionado. Para iniciar el proceso de extracción el usuario debe hacer click sobre el botón [Login to Facebook] para iniciar sesión y Luego click sobre el botón: [Fetch Data]

Figura N° 71: Configuración Consulta Extracción de Comentarios



Luego de hacer click sobre el botón Fetch Data, el proceso de extracción envía una petición a la API Graph, el cual retorna los datos retornados por la API, los cuales podemos visualizar en la ventana de visualización de nodos Figura N° 72.

Figura N° 72: Comentarios Retornados por a API Graph

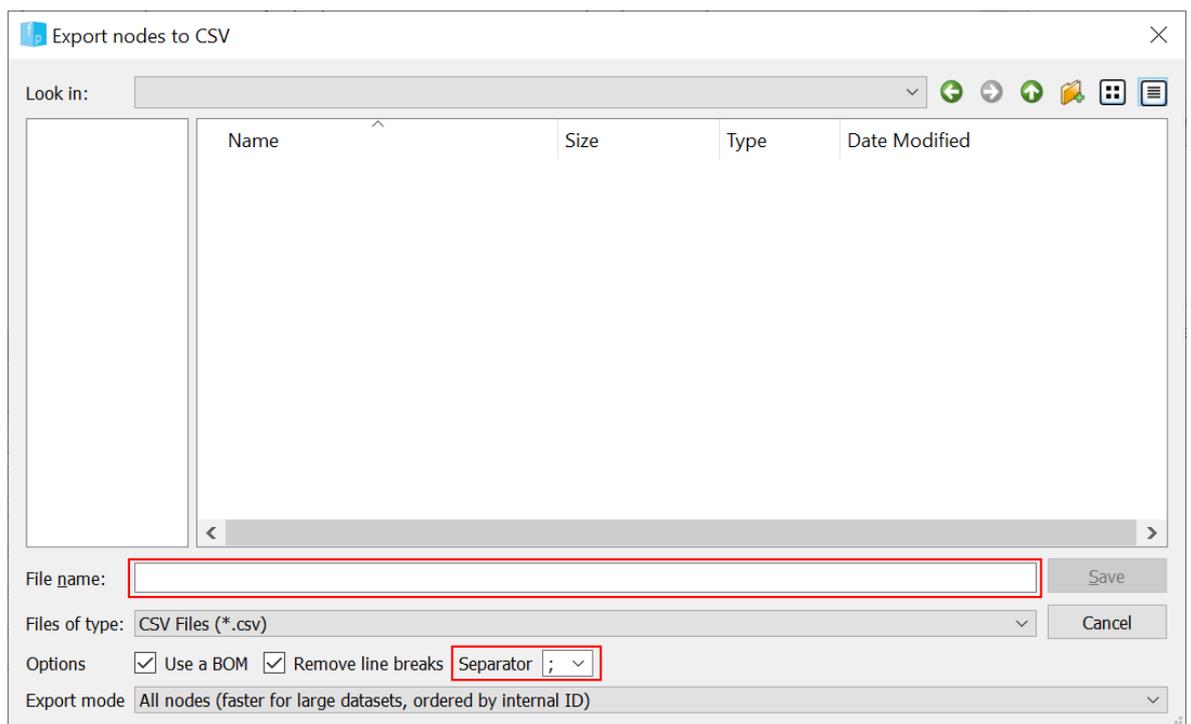
Object ID	Object Type	Object Key	Query Status	Query Time	Query Type	message	created_time
404953776356455	seed						
> 404953776356455_171491...	data	data.*	fetchd (200)	2021-07-04...	Facebook;/...	De lo que siempr...	2021-06-30T22...
> 404953776356455_171501...	data	data.*	fetchd (200)	2021-07-04...	Facebook;/...	¡Faltan pocos día...	2021-06-30T18...
404953776356455_171253...	data	data.*	fetchd (200)	2021-07-04...	Facebook;/...	¡Protege tu salud ...	2021-06-23T15...
404953776356455_170934...	data	data.*	fetchd (200)	2021-07-04...	Facebook;/...	¿Tienes consultas ...	2021-06-19T16...
404953776356455_170806...	data	data.*	fetchd (200)	2021-07-04...	Facebook;/...	Nuestros asesore...	2021-06-18T15...
404953776356455_170329...	data	data.*	fetchd (200)	2021-07-04...	Facebook;/...	¡No lo pienses m...	2021-06-10T21...
404953776356455_170158...	data	data.*	fetchd (200)	2021-07-04...	Facebook;/...	Gracias por siem...	2021-06-08T21...
404953776356455_169891...	data	data.*	fetchd (200)	2021-07-04...	Facebook;/...	Está en nosotros ...	2021-06-05T15...
404953776356455_169724...	data	data.*	fetchd (200)	2021-07-04...	Facebook;/...	Tenemos una am...	2021-06-02T20...
404953776356455_169347...	data	data.*	fetchd (200)	2021-07-04...	Facebook;/...	Conoce más ben...	2021-05-28T22...
> 404953776356455_168776...	data	data.*	fetchd (200)	2021-07-04...	Facebook;/...	Nuestros asesore...	2021-05-20T23...

### 5.7.1.2.5. Exportar nodos a archivos .csv

Para exportar los datos retornados por la API Graph a un archivo portable con extensión .csv, se debe hacer click sobre el botón [Export Data] en la barra de menús de la ventana principal.

Inmediatamente se abrirá la ventana (Figura N° 73): Export nodes to CSV en la cual se puede configurar el nombre del archivo y el carácter de separación de columnas.

Figura N° 73: Exportar Nodos a Archivo CSV



### **5.7.2. Conjunto de datos de entrenamiento**

Para iniciar el proceso de entrenamiento de los modelos de: Clasificación de Texto y de Análisis de Sentimiento, como se puede apreciar en la Figura N° 74 se ha creado el conjunto de datos de entrenamiento con las siguientes columnas y valores las cuales detallamos a continuación:

- Comentario: Son los comentarios expresados por usuarios en las redes sociales quienes reaccionan ante una publicación.
- Tema: relacionados al rubro de seguros: Soat, Vehicular, Seguro de Salud, Seguros de Vida, etc.
- Polaridad:
  - P: Positivo
  - N: Negativo
  - NEU: Neutro

Figura N° 74: Conjunto de Datos de Entrenamiento

comentario	tema	polaridad
:v una llanta hasta mi perro lo cambia xd	Vehicular	N
¿Soat para moto? Precio por favor	Soat	P
A cuanto vale el seguro	Información	P
A qué número debo llamar para SOAT por favor	Soat	P
Hay que ser bien gil para no saber cambiar una llanta, bueno a u	Vehicular	N
Aceite y agua se verifica diario..	Vehicular	NEU
Hola, quiero cancelar o cambiar el seguro por otro monto.	Información	N
Algún número para llamar mme interesa	Información	P
Alguna promoción para el seguro de autos???	Información	NEU
Aun tiene la promocion ???	Información	NEU
Auto antiguo cuenta?	Vehicular	NEU
Bota humo negro	Vehicular	N
Buenas noches, información para LADA NIVA 2016	Vehicular	P
Buenos dias, el precio de soat para moto cuanto cuesta por favo	Soat	P
Costo del SOAT particular para KIA PICANTO	Soat	P
Costo, soat y seguro	Soat	P
Inf..q coberturas cubre..100% reembolsable ???	Cobertura	NEU
Informes para el seguro de vida??	Vida	NEU

Fuente: Elaboración Propia (2020)

### 5.7.3. Clasificación de texto

Para poder asignar en una categoría o clase un comentario extraído desde las redes sociales en el paso de extracción de comentarios, vamos a evaluar y seleccionar el algoritmo de clasificación que tengan mejor desempeño o performance.

La clasificación o categorización de texto es el proceso de colocar un documento en una categoría o clase de acuerdo con las características propias del documento. En la minería de texto, la clasificación se refiere a la actividad de analizar o estudiar la colección de documentos de texto

preclasificados para derivar un modelo o función que se puede usar para agrupar otros documentos de clase que son desconocidos en una o más clases predefinidas.

Los datos extraídos desde las redes sociales serán cargados y procesados durante la clasificación de texto. Para la presente investigación los algoritmos de clasificación de texto que será evaluados son: Naive Bayes, Decision Tree y Support Vector Machine (SVM).

Antes de entrar en la etapa de procesamiento del algoritmo de clasificación de texto, los datos pasarán por la etapa de preprocesamiento, como son:

- Case folding
- Tokenización
- Filtrado
- Stemming (para eliminar los ruidos de datos)

#### **5.7.3.1. Selección del algoritmo de clasificación**

Para seleccionar el algoritmo de clasificación de texto debemos evaluar cada algoritmo en base a su performance. El método de evaluación que se utilizará será mediante la aplicación del esquema test con 5 composiciones de datos diferentes y también para cada una de las composiciones de prueba se realizará aplicando el método 10-fold cross-validation.

De acuerdo, a la investigación de los autores (Kurnia & Suharjito, 2018), en la clasificación hay varias formas de medir la performance de los algoritmos, como con son: la exactitud, precisión, recuperación y f-measure.

Los resultados de los test aplicados a los algoritmos de clasificación, usando 10-Fold Cross Validation, se muestran en las Tablas N° 11, 12 y 13:

Tabla N° 11: Resultados test algoritmo Naive Bayes

Iteration	Test Case				
	1st	2nd	3rd	4th	5th
K-1	73.076%	72.448%	73.426%	74.479%	72.033%
K-2	60.000%	65.979%	69.718%	73.936%	75.213%
K-3	71.428%	76.842%	80.141%	74.731%	70.689%
K-4	64.583%	70.967%	71.428%	76.086%	75.324%
K-5	65.217%	68.817%	71.942%	73.224%	74.458%
K-6	60.869%	77.173%	75.362%	73.770%	75.217%
K-7	59.090%	61.111%	78.676%	75.824%	71.491%
K-8	66.667%	79.545%	68.382%	75.274%	75.000%
K-9	61.904%	70.454%	77.037%	75.274%	74.561%
K-10	42.857%	70.454%	72.180%	74.274%	74.889%
<b>Average</b>	62.569%	71.379%	73.829%	74.787%	73.888%

Fuente: Business Intelligence Model to Analyze Social Media Information (Kurnia & Suharjito, 2018)

Tabla N° 12: Resultados test algoritmo Decision Tree

Iteration	Test Case				
	1st	2nd	3rd	4th	5th
K-1	52.941%	51.546%	54.545%	61.170%	57.203%
K-2	46.938%	46.391%	63.380%	59.893%	62.127%
K-3	51.020%	56.842%	54.285%	59.893%	56.896%
K-4	46.808%	48.936%	55.395%	59.893%	61.471%
K-5	51.111%	47.872%	56.521%	56.756%	62.173%
K-6	55.556%	58.064%	48.175%	55.434%	59.388%
K-7	45.454%	46.667%	57.664%	60.109%	55.021%
K-8	56.818%	53.409%	50.735%	56.830%	51.091%
K-9	36.363%	54.022%	66.911%	57.222%	54.385%
K-10	46.511%	62.068%	56.296%	49.444%	56.637%
<b>Average</b>	48.952%	52.582%	56.391%	57.664%	57.639%

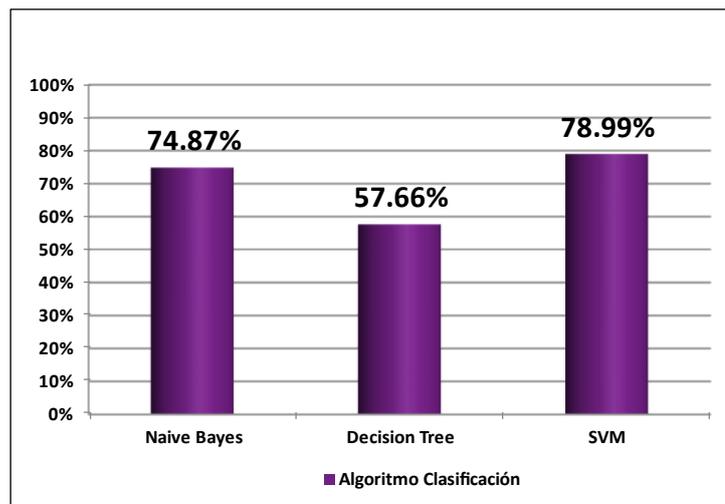
Fuente: Business Intelligence Model to Analyze Social Media Information (Kurnia & Suharjito, 2018)

Tabla N° 13: Resultados test algoritmo SVM

Iteration	Test Case				
	1st	2nd	3rd	4th	5th
<b>K-1</b>	76.470%	79.381%	81.560%	73.684%	77.966%
<b>K-2</b>	78.431%	74.736%	80.000%	75.661%	76.694%
<b>K-3</b>	81.250%	87.234%	79.285%	72.727%	74.678%
<b>K-4</b>	76.595%	74.468%	74.100%	77.419%	73.706%
<b>K-5</b>	71.739%	82.795%	78.417%	74.054%	74.347%
<b>K-6</b>	84.782%	77.173%	71.223%	80.874%	77.826%
<b>K-7</b>	84.090%	77.778%	77.697%	74.585%	77.092%
<b>K-8</b>	76.744%	73.333%	76.470%	76.243%	75.770%
<b>K-9</b>	86.046%	81.111%	78.676%	78.453%	75.330%
<b>K-10</b>	73.809%	71.264%	68.656%	77.900%	74.889%
<b>Average</b>	78.996%	77.927%	76.608%	76.160%	75.830%

Fuente: Business Intelligence Model to Analyze Social Media Information (Kurnia & Suharjito, 2018)

Figura N° 75: Comparación de Precisión algoritmos: Naive Bayes, Decision Tree y SVM



Fuente: Business Intelligence Model to Analyze Social Media Information (Kurnia & Suharjito, 2018)

La Figura N° 75 describe que el algoritmo SVM tiene una mejor precisión que los algoritmos: Naive Bayes y Decision Tree, por lo que, si los ordenamos según el valor de precisión, el algoritmo SVM se ubica en el primer lugar con una precisión del 78,99%, en segundo lugar, se ubica el

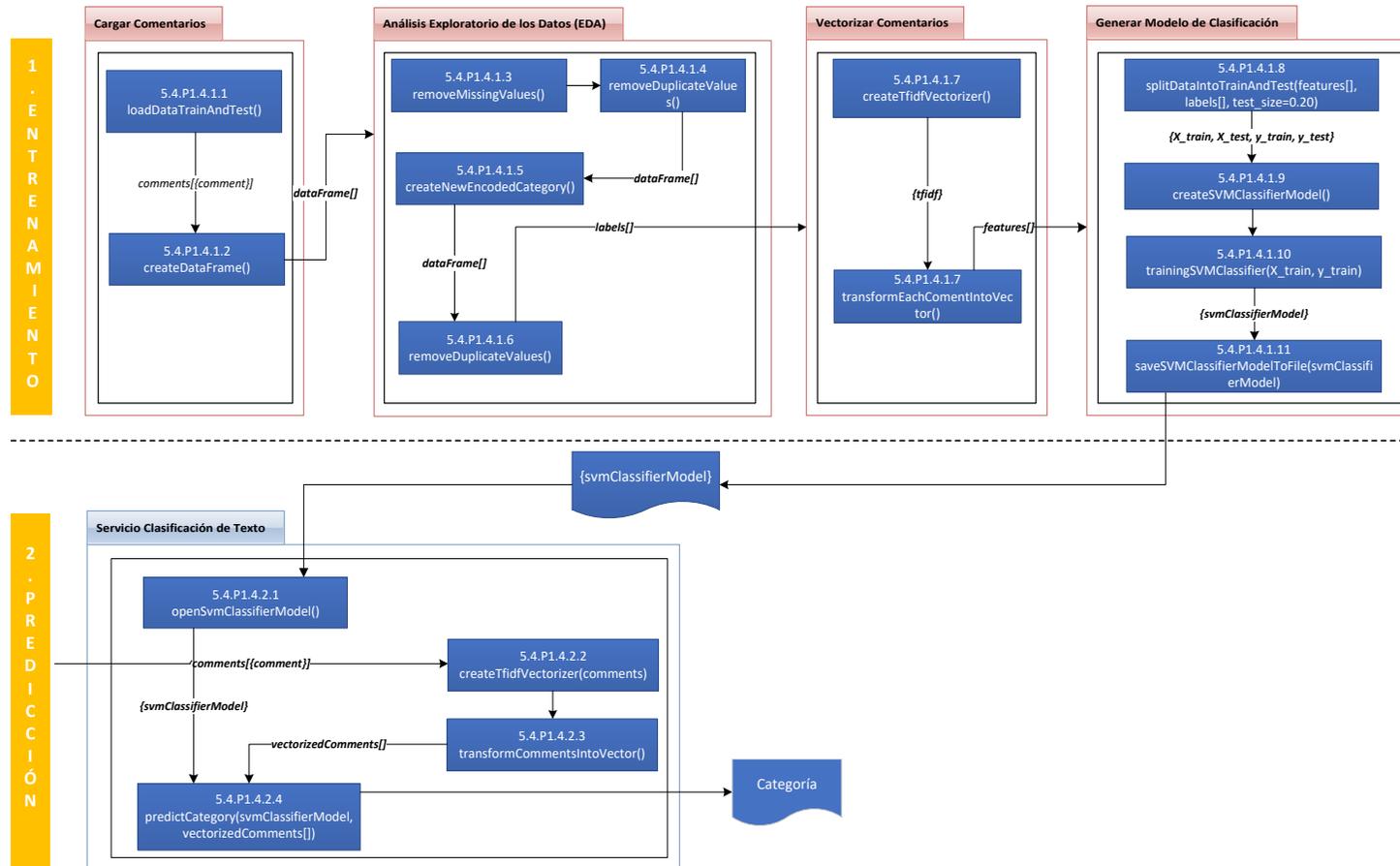
algoritmo Naive Bayes con una precisión del 74,78% y finalmente el algoritmo Decision Tree que tiene una precisión del 57,66%.

Luego de realizar esta comparación se concluye que el algoritmo de clasificación de texto que se utilizará en la presente investigación será: Support Vector Machines (SVM).

### 5.7.3.2. Modelo de clasificación de texto

Después de realizar la evaluación de los algoritmos de clasificación, se procede a diseñar el modelo de clasificación de texto (Figura N° 76) utilizando el algoritmo de clasificación SVM el cual permitirá asignar un comentario no categorizado en una categoría existente, utilizando el siguiente modelo de clasificación de texto.

Figura N° 76: Modelo de Clasificación de Texto



Fuente: Elaboración Propia (2021)

#### 5.7.4. Análisis de sentimiento

El análisis de sentimiento también conocido como minería de opinión, es un campo de la investigación que analiza las opiniones, sentimientos, evaluaciones, valoraciones, actitudes y emociones de las personas que son expresadas través de comentarios, revisión de productos, etc.

##### 5.7.4.1. Librerías del modelo

Las librerías utilizadas para la construcción del modelo de Análisis de Sentimiento son: Tensorflow y Keras (Figura 77):

Figura N° 77: Librerías del Modelo Análisis de Sentimiento

```
import pandas as pd
import numpy as np
from tensorflow.keras.preprocessing.text import one_hot
from tensorflow.keras.preprocessing.sequence import pad_sequences
from tensorflow.keras.models import Sequential, load_model, save_model, Model
from tensorflow.keras.layers import Dense
from tensorflow.keras.layers import Flatten
from tensorflow.keras.layers import Conv1D
from tensorflow.keras.layers import MaxPool1D
from keras.utils import np_utils
from tensorflow.keras.layers import Embedding
from tensorflow.keras.preprocessing.text import Tokenizer

from sklearn.model_selection import train_test_split
import tempfile
from io import StringIO
import pickle
```

### 5.7.4.2. Preparación de los datos

En tareas de procesamiento de lenguaje natural, el preprocesamiento de datos es parte importante del trabajo. Debido a que los algoritmos no entienden los textos, se hace necesario convertir los comentarios en números que puedan ser entendidos por los algoritmos (Figura N° 78 y 79).

Figura N° 78: Carga del Conjunto de Datos de Entrenamiento

```
url = "C:\\\\.\\datasets\\"
df = pd.read_csv(url+'rows_seguros.csv', engine='python')
print(df.head())
```

	message	created_time	Product	polarity
0	Antony David Diaz Carrasco mira	2021-02-28T17:58:24+0000	Información	P
1	:v una llanta hasta mi perro lo cambia xd	2019-05-19T03:07:36+0000	Vehicular	N
2	????	2019-11-24T02:03:45+0000	Información	NEU
3	¿Cómo lo hago, dónde dice? ????	2020-08-06T19:09:36+0000	Información	NEU
4	¿Soat para moto? Precio por favor	2020-02-01T18:46:49+0000	Soat	NEU

Figura N° 79: Procesamiento del Conjunto de Datos de Entrenamiento

```
#Conjunto de Entrenamiento
x_message = df['message'].to_list()
print(x_message[:4])

df['polarity_num'] = df['polarity'].apply(lambda x: ['N', 'P', 'NEU'].index(x))

Y = df['polarity_num'].values
print(Y[:10])
```

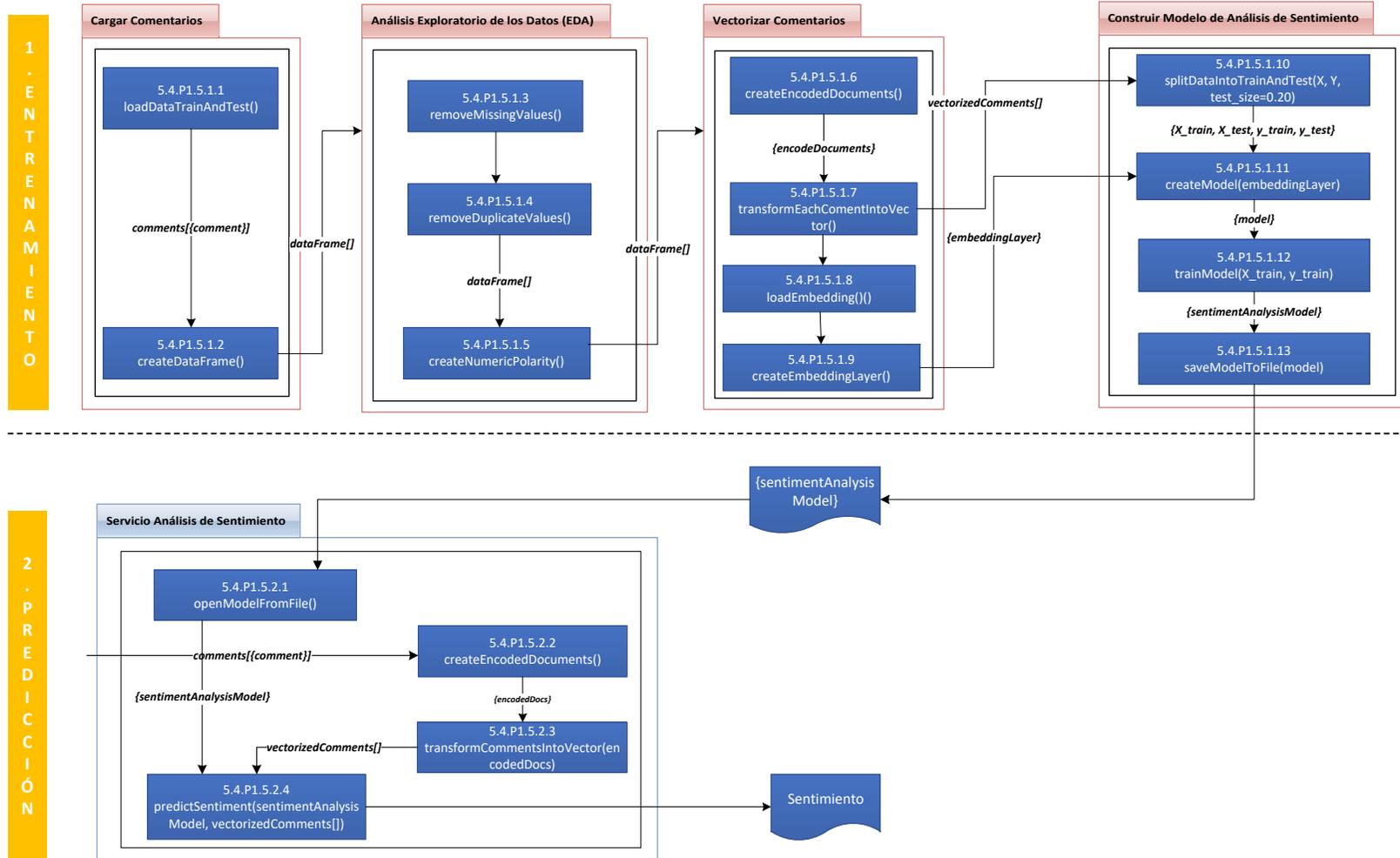
```
[' Antony David Diaz Carrasco mira', ':v una llanta hasta mi perro lo cambia xd', '????', '¿Cómo lo hago, dónde dice? ????']
```

```
[1 0 2 2 2 0 3 2 2 2]
```

### 5.7.4.3. Modelo de análisis de sentimiento

En la Figura N° 80, se podrá encontrar el modelo de análisis de sentimiento propuesta en la presente investigación, en la cual permitirá transformar las emociones expresadas en un comentario en información objetiva denominada sentimiento: Positivo, negativo o neutro.

Figura N° 80: Modelo de Análisis de Sentimiento



Fuente: Elaboración Propia (2021)

### 5.7.5. Construcción del almacén de datos

El diseño y construcción del almacén de datos se realizará utilizando la metodología Kimball la cual propone 4 fases durante el diseño del almacén de datos, es decir, seleccionar el proceso comercial, declarar la granularidad, identificar las dimensiones e identificar los hechos.

Para la construcción del repositorio de datos de la solución se utilizará el motor de base datos: Microsoft SQL Server 2016. Como se puede apreciar en la Tabla N° 14, cada uno de los componentes de la arquitectura de la solución propuesta, requiere de una base datos para almacenar lo datos en cada una de las fases: Data Staging Area (Tabla N° 15), Operational Data Store (Tabla N° 16) y Datamart.

Tabla N° 14: Bases de datos de la solución

	<b>Nombre de Servidor</b>	<b>Nombre Base Datos</b>
Data Staging Area (STG)		
Extracción Redes Sociales	PE-LIM-DBSE-01P	MBIS-STG-01
Limpieza de datos	PE-LIM-DBSE-01P	MBIS-STG-02
Operational Data Store (ODS)	PE-LIM-DBSE-01P	MBIS-ODS-01
Datamart (DWH)	PE-LIM-DBSE-01P	MBIS-DWH-01

Fuente: Elaboración Propia (2019)

### 5.7.5.1. Data Staging Area (STG)

Tabla N° 15: Características data staging area

	Características
<b>Motor Base Datos</b>	SQL Server 2016
<b>Nombre Servidor</b>	PE-LIM-DBSE-01P
<b>Nombre Base Datos</b>	MBIS-STG-01
<b>Sistema Operativo</b>	Windows Server 2016
<b>Nombre de Archivo</b>	
<b>Datos Lógico</b>	MBIS-STG-01_dat MBIS-STG-01_log
<b>Ruta Física</b>	
<b>Datos Lógico</b>	E:\Data\SQL2K16\MBIS-STG-01\DATA\MBIS-STG-01.mdf F:\Data\SQL2K16\MBIS-STG-01\LOG\MBIS-STG-01_log.ldf

Fuente: Elaboración Propia (2019)

### 5.7.5.2. Operational Data Store (ODS)

Tabla N° 16: Base datos operational data store (ODS)

	Características
<b>Motor Base Datos</b>	SQL Server 2016
<b>Nombre Servidor</b>	LIM1-DBSE-01P
<b>Nombre Base Datos</b>	MBIS-ODS-01
<b>Sistema Operativo</b>	Windows Server 2016
<b>Nombre de Archivo</b>	
<b>Datos Lógico</b>	MBIS-ODS-01_dat MBIS-ODS-01_log
<b>Ruta Física</b>	
<b>Datos Lógico</b>	E:\Data\SQL2K16\MBIS-ODS-01\DATA\MBIS-ODS-01.mdf F:\Data\SQL2K16\MBIS-ODS-01\LOG\MBIS-ODS-01_log.ldf

Fuente: Elaboración Propia (2019)

### 5.7.5.3. Datamart area (DWH)

Un paso fundamental en el desarrollo de la solución será la creación de la base de datos para almacenar el datamart del modelo, ya que nos permitirá almacenar los datos del modelo en una base datos relacional, la cual servirá luego como fuente de datos de la base datos dimensional (OLAP). En la Tabla N° 17 se puede encontrar las características que debe tener esta base de datos.

Tabla N° 17: Características base datos

	Características
<b>Motor Base Datos</b>	SQL Server 2016
<b>Nombre Servidor</b>	PE-LIM-DBSE-01P
<b>Nombre Base Datos</b>	MBIS-DWH-01
<b>Sistema Operativo</b>	Windows Server 2016
<b>Nombre de Archivo</b>	
<b>Datos Lógico</b>	MBIS-DWH-01_dat MBIS-DWH-01_log
<b>Ruta Física</b>	
<b>Datos Lógico</b>	E:\Data\SQL2K16\MBIS-DWH-01\DATA\MBIS-DWH-01.mdf F:\Data\SQL2K16\MBIS-DWH-01\LOG\MBIS-DWH-01_log.ldf

Fuente: Elaboración Propia (2019)

#### 5.7.5.4. Conexión al servidor de base datos

Para iniciar con el proceso de creación de las bases de datos del modelo de inteligencia de negocios, se utilizará la herramienta cliente: SQL Server Management Studio (SSMS).

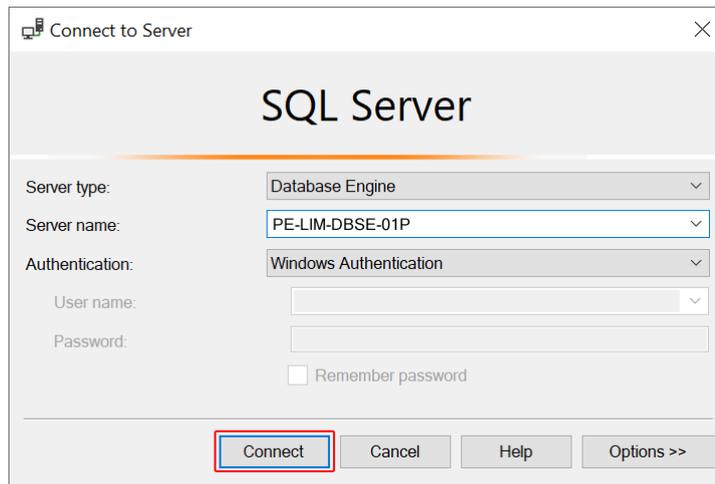
Para abrir una nueva conexión al servidor de base datos, desde la herramienta cliente, SQL Server Management Studio (SSMS), dar click sobre el botón nueva conexión. Como se puede ver en la Figura N° 81, inmediatamente se abrirá una ventana que le permitirá al usuario ingresar los parámetros de conexión:

- Tipo de Servidor: Motor de Base Datos
- Nombre de Servidor: LIM1-DBSE-01P
- Autenticación: Autenticación de Windows

Luego de ingresar los parámetros de conexión, dar click sobre el botón conectar e inmediatamente se abrirá la herramienta de consulta en la

cual se podrán escribir los comandos de creación de base datos y de las tablas.

Figura N° 81: Conexión al Servidor de Base Datos



#### 5.7.5.5. Creación de carpetas del modelo

Un paso fundamental para la creación de las bases de datos es la creación de las carpetas, en donde se almacenarán en forma física los archivos de datos y los archivos lógicos de cada base de datos, se crearán la estructura de carpetas en el servidor de base datos de acuerdo con la Tabla N° 18.

Tabla N° 18: Carpetas del modelo

Base Datos	Tipo de Archivo	Ruta Física	Nombre de Archivo
MBIS-STG-01	Datos	E:\Data\SQL2K16\MBIS-STG-01\DATA	MBIS-STG-01.mdf
MBIS-STG-01	Lógico	F:\Data\SQL2K16\MBIS-STG-01\LOG	MBIS-STG-01_log.ldf
MBIS-ODS-01	Datos	E:\Data\SQL2K16\MBIS-ODS-01\DATA	MBIS-ODS-01.mdf
MBIS-ODS-01	Lógico	F:\Data\SQL2K16\MBIS-ODS-01\LOG	MBIS-ODS-01_log.ldf
MBIS-DWH-01	Datos	E:\Data\SQL2K16\MBIS-DWH-01\DATA	MBIS-DWH-01.mdf
MBIS-DWH-01	Lógico	F:\Data\SQL2K16\MBIS-DWH-01\LOG	MBIS-DWH-01_log.ldf

Fuente: Elaboración Propia (2020)

Ingresar al servidor de base de datos y crear la estructura de carpetas para cada una de las bases de datos del modelo de acuerdo con las Figuras: N° 82, 83 y 84:

Figura N° 82: Estructura de carpetas base datos: MBIS-STG-01

his PC > Local Disk (M:) > Data > SQL2K16 > MBIS-STG-01 > DATA		
Name	Date modified	Type
 MBIS-STG-01.mdf	26/06/2022 05:22	SQL Server Database Primary Data File
his PC > Local Disk (M:) > Log > SQL2K16 > MBIS-STG-01 > LOG		
Name	Date modified	Type
 MBIS-STG-01_log.ldf	26/06/2022 05:22	SQL Server Database Transaction Log File

Figura N° 83: Estructura de carpetas base datos: MBIS-ODS-01

his PC > Local Disk (M:) > Data > SQL2K16 > MBIS-ODS-01 > DATA		
Name	Date modified	Type
 MBIS-ODS-01.mdf	5/07/2022 16:48	SQL Server Database Primary Data File
his PC > Local Disk (M:) > Log > SQL2K16 > MBIS-ODS-01 > LOG		
Name	Date modified	Type
 MBIS-ODS-01_log.ldf	5/07/2022 16:48	SQL Server Database Transaction Log File

Figura N° 84: Estructura de carpetas base datos: MBIS-DWH-01

his PC > Local Disk (M:) > Data > SQL2K16 > MBIS-DWH-01 > DATA		
Name	Date modified	Type
 MBIS-DWH-01.mdf	5/07/2022 16:57	SQL Server Database Primary Data File
his PC > Local Disk (M:) > Log > SQL2K16 > MBIS-DWH-01 > LOG		
Name	Date modified	Type
 MBIS-DWH-01_log.ldf	5/07/2022 16:57	SQL Server Database Transaction Log File

### 5.7.5.6. Creación bases de datos del modelo

Para iniciar con el proceso de creación de las bases de datos del modelo, se debe ejecutar la herramienta cliente SQL Server Management Studio, luego abrir y ejecutar los siguientes scripts de creación de bases de datos: Figura N° 85 Data Stage, Figura N° 86 ODS y Figura N° 87 Datamart.

Figura N° 85: Script creación de base datos [MBIS-STG-01]

```
CREATE DATABASE [MBIS-STG-01] ON PRIMARY
( NAME = N'MBIS-STG-01_dat',
  FILENAME = N'M:\Data\SQL2K16\MBIS-STG-01\DATA\MBIS-STG-01.mdf' , SIZE
= 51200KB , FILEGROWTH = 1024KB )
LOG ON
( NAME = N'MBIS-STG-01_log',
  FILENAME = N'M:\Log\SQL2K16\MBIS-STG-01\LOG\BIS-STG-01_log.ldf' , SIZE
= 1024KB , FILEGROWTH = 10%)
GO
```

Figura N° 86: Script creación de base datos [MBIS-ODS-01]

```
CREATE DATABASE [MBIS-ODS-01] ON PRIMARY
( NAME = N'MBIS-ODS-01_dat',
  FILENAME = N'M:\Data\SQL2K16\MBIS-ODS-01\DATA\MBIS-ODS-01.mdf' , SIZE
= 51200KB , FILEGROWTH = 1024KB )
LOG ON
( NAME = N'MBIS-ODS-01_log',
  FILENAME = N'M:\Log\SQL2K16\MBIS-ODS-01\LOG\BIS-ODS-01_log.ldf' , SIZE
= 1024KB , FILEGROWTH = 10%)
GO
```

Figura N° 87: Script creación de base datos [MBIS-DWH-01]

```
CREATE DATABASE [MBIS-DWH-01] ON PRIMARY
( NAME = N'MBIS-DWH-01_dat',
  FILENAME = N'M:\Data\SQL2K16\MBIS-DWH-01\DATA\MBIS-DWH-01.mdf' , SIZE
= 51200KB , FILEGROWTH = 1024KB )
LOG ON
( NAME = N'MBIS-DWH-01_log',
  FILENAME = N'M:\Log\SQL2K16\MBIS-DWH-01\LOG\MBIS-DWH-01_log.ldf' ,
SIZE = 1024KB , FILEGROWTH = 10%)
GO
```

## 5.7.5.7. Creación tablas del modelo

### 5.7.5.7.1. Tablas data staging area

Para realizar los trabajos de carga, limpieza, estandarización de los datos, se necesita del apoyo de tablas que servirán como repositorio de datos temporales, es decir serán volátiles y serán eliminadas cada vez que se ejecuta un nuevo proceso de carga. Estas tablas se encuentran en el Data Staging Area, para crear las tablas se debe ejecutar los siguientes scripts de las Figuras 88, 89, 90, 91, 92 y 93.

Luego podremos observar el diagrama de la base datos data staging area en la Figura N° 94.

Figura N° 88: Script tabla FACEBOOK\_000

```
USE [MBIS-STG-01]
GO

CREATE TABLE dbo.STG_FACEBOOK_000
(
    [Col001] varchar(200) null,           --level
    [Col002] varchar(200) NULL,         --id
    [Col003] varchar(200) null,         --parent_id
    [Col004] varchar(200) NULL,         --object_id
    [Col005] varchar(200) NULL,         --object_type
    [Col006] varchar(200) NULL,         --object_key
    [Col007] varchar(200) NULL,         --query_status
    [Col008] varchar(200) NULL,         --query_time
    [Col009] varchar(200) NULL,         --query_type
    [Col010] ntext NULL,                --message
    [Col011] varchar(200) NULL,         --created_time
    [Col012] varchar(200) NULL,         --updated_time
    [Col013] varchar(200) NULL,         --error.message
    [Col014] varchar(200) NULL,         --from.name
    [Col015] varchar(200) NULL,         --from.id
)
GO
```

Figura N° 89: Script tabla FACEBOOK\_001

```
USE [MBIS-STG-01]
GO

CREATE TABLE dbo.STG_FACEBOOK_001 (
    [level]          int null,
    [id] int NULL,
    [parent_id] int null,
    [object_id] varchar(200) NULL,
    [object_type] varchar(200) NULL,
    [object_key] varchar(200) NULL,
    [query_status] varchar(200) NULL,
    [query_time] datetime,
    [query_type] varchar(200) NULL,
    [message] ntext NULL,
    [created_time] datetime NULL,
    [updated_time] datetime NULL,
    [error.message] varchar(200) NULL,
    [from.name] varchar(200) NULL,
    [from.id] varchar(200) NULL
)
GO
```

Figura N° 90: Script tabla FACEBOOK\_002

```
USE [MBIS-STG-01]
GO

CREATE TABLE dbo.STG_FACEBOOK_002 (
    [level]          int null,
    [id] int NULL,
    [parent_id] int null,
    [object_id] varchar(200) NULL,
    [object_type] varchar(200) NULL,
    [object_key] varchar(200) NULL,
    [query_status] varchar(200) NULL,
    [query_time] datetime,
    [query_type] varchar(200) NULL,
    [message] varchar(MAX) NULL,
    [created_time] datetime NULL,
    [updated_time] datetime NULL,
    [error.message] varchar(200) NULL,
    [from.name] varchar(200) NULL,
    [from.id] varchar(200) NULL,
    DescripcionTema varchar(100) NULL CONSTRAINT
DF_STG_FACEBOOK_002_DescripcionTema DEFAULT 'N/A'
)
GO
```

Figura N° 91: Script tabla FACEBOOK\_003

```
USE [MBIS-STG-01]
GO

CREATE TABLE dbo.STG_FACEBOOK_003 (
    [level] int null,
    [id] int NULL,
    [parent_id] int null,
    [object_id] varchar(200) NULL,
    [object_type] varchar(200) NULL,
    [object_key] varchar(200) NULL,
    [query_status] varchar(200) NULL,
    [query_time] datetime,
    [query_type] varchar(200) NULL,
    [message] varchar(MAX) NULL,
    [created_time] datetime NULL,
    [updated_time] datetime NULL,
    [error.message] varchar(200) NULL,
    [from.name] varchar(200) NULL,
    [from.id] varchar(200) NULL,
    DescripcionTema varchar(100) NULL,
    DescripcionSentimiento varchar(100) NULL CONSTRAINT
DF_STG_FACEBOOK_003_DescripcionSentimiento DEFAULT 'N/A'
)
GO
```

Figura N° 92: Script tabla FACEBOOK\_004

```
USE [MBIS-STG-01]
GO

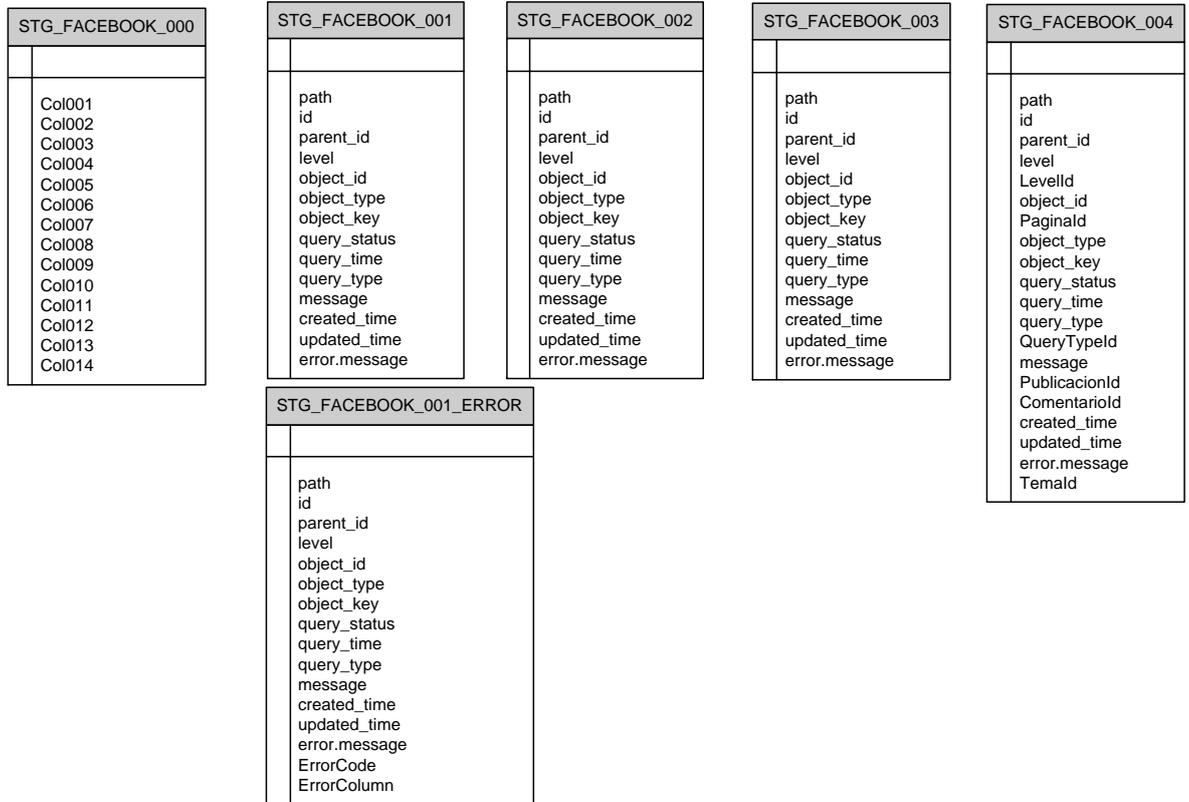
CREATE TABLE dbo.STG_FACEBOOK_004 (
    [level]          int null,
    [LevelId] int NULL,
    -----
    [id] int NULL,
    [parent_id] int null,
    -----
    [object_id] varchar(200) NULL,
    [PaginaId] int NULL,
    -----
    [object_type] varchar(200) NULL,
    [object_key] varchar(200) NULL,
    [query_status] varchar(200) NULL,
    [query_time] datetime,
    -----
    [query_type] varchar(200) NULL,
    [QueryTypeId] int NULL,
    -----
    [message] varchar(MAX) NULL,
    [PublicacionId] int NULL,
    [ComentarioId] int NULL,
    -----
    [created_time] datetime NULL,
    [CreatedTimeId] int NULL,
    [updated_time] datetime NULL,
    [UpdatedTimeId] int NULL,
    [error.message] varchar(200) NULL,
    [from.name] varchar(200) NULL,
    [from.id] varchar(200) NULL,
    ----- Tema
    [DescripcionTema] varchar(100) NULL,
    [TemaId] smallint NULL,
    ----- Sentimiento
    [DescripcionSentimiento] varchar(100) NULL,
    [SentimientoId] smallint NULL
)
GO
```

Figura N° 93: Script tabla FACEBOOK\_001\_ERROR

```
USE [MBIS-STG-01]
GO

CREATE TABLE dbo.STG_FACEBOOK_001_ERROR (
    [path] varchar(200) NULL,
    [id] int NULL,
    [parent_id] int null,
    [level] int null,
    [object_id] varchar(200) NULL,
    [object_type] varchar(200) NULL,
    [object_key] varchar(200) NULL,
    [query_status] varchar(200) NULL,
    [query_time] datetime, --Datetime
    [query_type] varchar(200) NULL,
    [message] ntext NULL,
    [created_time] VARCHAR(200) NULL, --Datetime
    [updated_time] VARCHAR(200) NULL, --Datetime
    [error.message] varchar(200) NULL,
    ErrorCode int NULL,
    ErrorColumn int NULL,
    [_created_time] VARCHAR(200) NULL
)
GO
```

Figura N° 94: Diagrama de tablas Data Staging Area



Fuente: Elaboración Propia (2020)

### 5.7.5.7.2. Tablas Data Warehouse area

Para la creación de la tabla de hechos se debe ejecutar el script de la Figura N° 95, luego para la creación de las dimensiones se deben ejecutar los scripts de la Figuras N° 96, 97, 98, 99, 100, 101 y finalmente para la creación de restricciones de se debe ejecutar el script de la Figura N° 102.

Figura N° 95: Script tabla de hechos

```
USE [MBIS-DWH-01]
GO

CREATE TABLE [dbo].[TMFACT_FACEBOOK]
(
    [RowId] int NOT NULL IDENTITY (1,1),
    [PaginaId] smallint NOT NULL,
    [LevelId] int NOT NULL,
    [QueryTypeId] int NOT NULL,
    [PublicacionId] int NOT NULL,
    [ComentarioId] int NOT NULL,
    [TemaId] smallint NOT NULL,
    [SentimientoId] smallint NOT NULL,
    [CreatedTimeId] int NOT NULL,
    [UpdatedTimeId] int NULL,
    [cant_comments] int NULL,
    [cant_replies] int NULL,
    [positivo] smallint NULL,
    [negativo] smallint NULL,
    [neutro] smallint NULL,
    CONSTRAINT [PK_TMFACT_FACEBOOK_RowId] PRIMARY KEY CLUSTERED
(
    [RowId] ASC
) ON [PRIMARY]
) ON [PRIMARY]
GO
```

Figura N° 96: Script tabla de dimensión: DIM\_FECHA

```
USE [MBIS-DWH-01]
GO

CREATE TABLE [dbo].[DIM_FECHA]
(
    [FechaID] smallint IDENTITY(1,1) NOT NULL,
    [FechaCalendario] datetime NULL,
    [Dia] tinyint NULL,
    [Mes] tinyint NULL,
    [Año] tinyint NULL,
    [Semana] tinyint NULL,
    [Trimestre] tinyint NULL,
    [Semestre] tinyint NULL,
    [NombreMes] varchar (12) NULL
    CONSTRAINT [PK_DIM_FECHA_FechaID] PRIMARY KEY CLUSTERED
(
    [FechaID] ASC
)
ON [PRIMARY],
CONSTRAINT [AK_DIM_FECHA_FechaCalendario] UNIQUE NONCLUSTERED
(
    [FechaCalendario] ASC
)
ON [PRIMARY]
) ON [PRIMARY]
GO
```

Figura N° 97: Script tabla de dimensión: DIM\_TEMA

```
USE [MBIS-DWH-01]
GO

CREATE TABLE dbo.[DIM_TEMA]
(
    [TemaID] smallint IDENTITY (1, 1) NOT NULL ,
    [CodigoTema] varchar (10) NOT NULL,
    [DescripcionTema] varchar (100)NOT NULL,
    CONSTRAINT [PK_DIM_TEMA_TemaID] PRIMARY KEY CLUSTERED
    (
        [TemaID] ASC
    )
    ON [PRIMARY]
) ON [PRIMARY]
```

Figura N° 98: Script tabla de dimensión: DIM\_SENTIMIENTO

```
USE [MBIS-DWH-01]
GO

CREATE TABLE dbo.[DIM_SENTIMIENTO]
(
    [SentimientoId] smallint IDENTITY (1, 1) NOT NULL ,
    [CodigoSentimiento] varchar (20) NULL,
    [DescripcionSentimiento] varchar (100)NOT NULL,
    CONSTRAINT [PK_DIM_SENTIMIENTO_SentimientoId] PRIMARY KEY CLUSTERED
    (
        [SentimientoId] ASC
    )
    ON [PRIMARY]
) ON [PRIMARY]
```

Figura N° 99: Script tabla de dimensión: DIM\_PAGINA

```
USE [MBIS-DWH-01]
GO

CREATE TABLE dbo.[DIM_PAGINA]
(
    [PaginaID] smallint IDENTITY (1, 1) NOT NULL ,
    [NombrePagina] varchar (100) NOT NULL,
    [Categoria] varchar (255) NULL,
    [Descripcion] varchar (max) NULL,
    [URL] varchar (200) NULL,
    [Ubicacion] varchar (200) NULL,
    [SitioWeb] varchar (200) NULL,
    [MeGusta] int NULL,
    [Fotos] int NULL,
    [FechaCreacion] datetime NULL,
    CONSTRAINT [PK_DIM_PAGINA_PaginaID] PRIMARY KEY CLUSTERED
    (
        [PaginaID] ASC
    )
    ON [PRIMARY]
) ON [PRIMARY]
GO
```

Figura N° 100: Script tabla de dimensión: DIM\_PUBLICACION

```
USE [MBIS-DWH-01]
GO

CREATE TABLE dbo.[DIM_PUBLICACION]
(
    [PublicacionId] int IDENTITY (1, 1) NOT NULL,
    [ObjectId] varchar (11) NOT NULL,
    [ParentId] int NULL,
    [Id] int NULL,
    [Publicacion] varchar (8000) NULL
CONSTRAINT [PK_DIM_PUBLICACION_PublicacionId] PRIMARY KEY CLUSTERED
(
    [PublicacionId] ASC
)
ON [PRIMARY]
) ON [PRIMARY]
GO
```

Figura N° 101: Script tabla de dimensión: DIM\_COMENTARIO

```
CREATE TABLE dbo.[DIM_COMENTARIO]
(
    [ComentarioId] int IDENTITY (1, 1) NOT NULL,
    [ObjectId] varchar (11) NOT NULL,
    [ParentId] int NULL,
    [Id] int NULL,
    [Comentario] varchar (8000) NULL
CONSTRAINT [PK_DIM_COMENTARIO_ComentarioId] PRIMARY KEY CLUSTERED
(
    [ComentarioId] ASC
)
ON [PRIMARY],
CONSTRAINT [AK_DIM_COMENTARIO_Registro] UNIQUE NONCLUSTERED
(
    [ObjectId],[ParentId],[Id]
)
ON [PRIMARY]
) ON [PRIMARY]
```

Figura N° 102: Script restricciones base de datos

```
ALTER TABLE [dbo].[TMFACT_FACEBOOK] WITH CHECK ADD
    CONSTRAINT [FK_TMFACT_FACEBOOK_DIM_FECHA] FOREIGN KEY
    (
        [CreatedTimeId]
    )
REFERENCES [dbo].[DIM_FECHA] ([FechaId])
GO

ALTER TABLE [dbo].[TMFACT_FACEBOOK] WITH CHECK ADD
    CONSTRAINT [FK_TMFACT_FACEBOOK_DIM_TEMA] FOREIGN KEY
    (
        [TemaId]
    )
REFERENCES [dbo].[DIM_TEMA] ([TemaId])
GO

ALTER TABLE [dbo].[TMFACT_FACEBOOK] WITH CHECK ADD
    CONSTRAINT [FK_TMFACT_FACEBOOK_DIM_SENTIMIENTO] FOREIGN KEY
    (
        [SentimientoId]
    )
REFERENCES [dbo].[DIM_SENTIMIENTO] ([SentimientoId])
GO

ALTER TABLE [dbo].[TMFACT_FACEBOOK]
    WITH CHECK ADD
    CONSTRAINT [FK_TMFACT_FACEBOOK_DIM_PAGINA] FOREIGN KEY
    (
        [PaginaId]
    )
REFERENCES [dbo].[DIM_PAGINA] ([PaginaId])
GO

ALTER TABLE [dbo].[TMFACT_FACEBOOK] WITH CHECK ADD
    CONSTRAINT [FK_TMFACT_FACEBOOK_DIM_PUBLICACION] FOREIGN KEY
    (
        [PublicacionId]
    )
REFERENCES [dbo].[DIM_PUBLICACION] ([PublicacionId])
GO

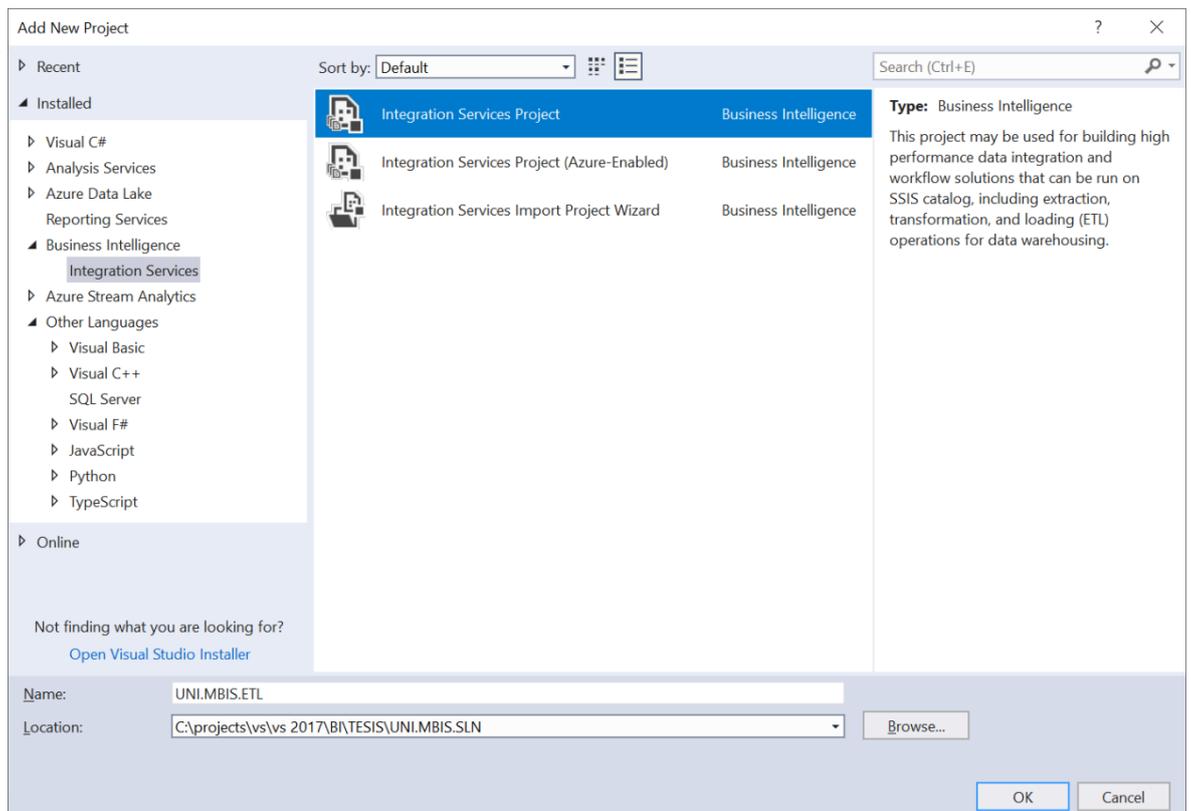
ALTER TABLE [dbo].[TMFACT_FACEBOOK] WITH CHECK ADD
    CONSTRAINT [FK_TMFACT_FACEBOOK_DIM_COMENTARIO] FOREIGN KEY
```

### 5.7.6. Creación del proyecto Integration Services (SSIS)

Para iniciar la creación de un proyecto Integration Services, se debe abrir la ventana de dialogo, utilizando la opción Archivo del menú principal y luego seleccionar Crear nuevo proyecto.

En la Figura N° 103, se puede visualizar la ventana de dialogo Crear nuevo proyecto, ubicar en la lista de categorías, la categoría: Business Intelligence, luego seleccionar la plantilla de proyecto: Integration Services, finalmente dar el nombre al proyecto: WTW.MBIS.ETL.

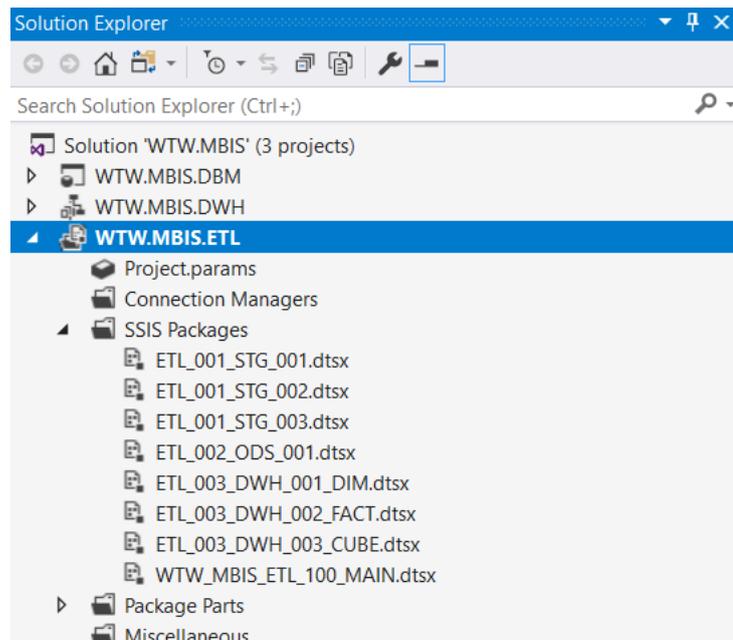
Figura N° 103: Selección de plantilla de proyecto Integration Services



Como se puede apreciar en la Figura N° 104, la herramienta SQL Server Data Tools creó el proyecto WTW.MBIS.ETL de acuerdo con la plantilla de

proyecto de Integration Services, el cual está compuesto de las siguientes carpetas de trabajo.

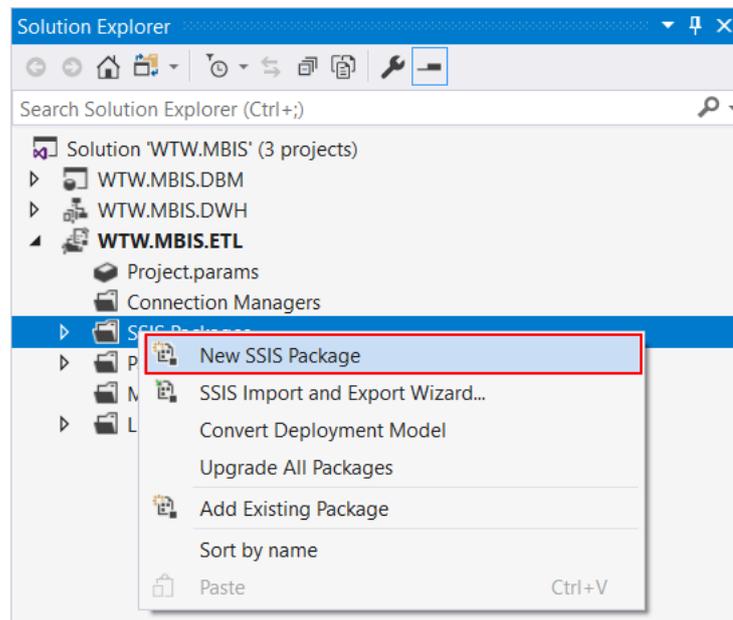
Figura N° 104: Carpetas plantilla de proyecto Integration Services



#### 5.7.6.1. Agregar paquete de Integration Services al proyecto

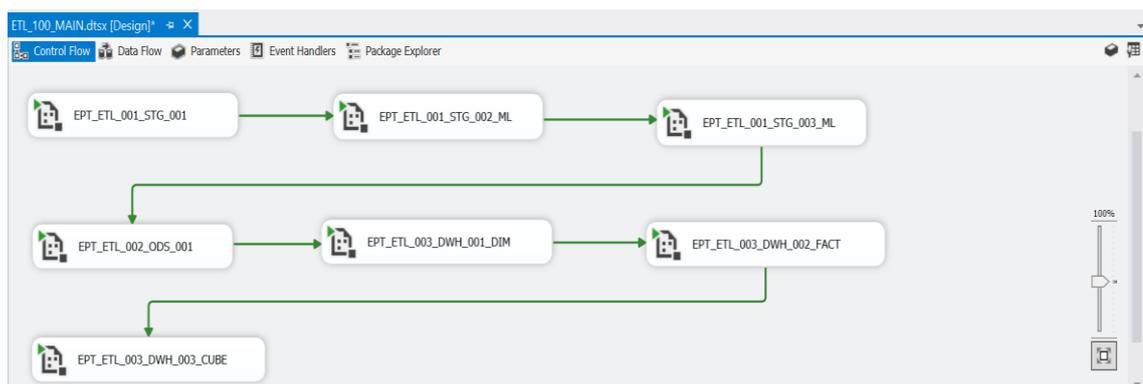
El siguiente paso será el agregar un nuevo paquete de Integration Services al proyecto. En la ventana: [Solution Explorer] ubicar la carpeta: [SSIS Packages] y dar click derecho sobre esta carpeta, inmediatamente se abrirá el menú contextual en el cual se debe seleccionar la opción: [New SSIS Packages...] (Figura N° 105)

Figura N° 105: Agregar nuevo paquete Integration Services



En la Figura N° 106 podemos visualizar el diseñador de paquetes, en la pestaña [Control Flow] encontramos los diferentes componentes de carga, transformación y procesamiento del cubo.

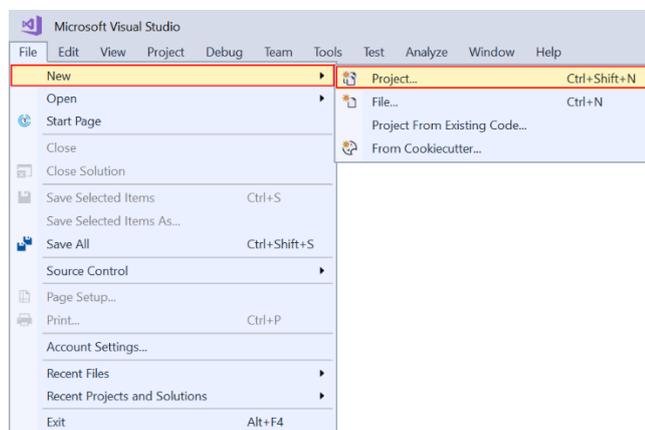
Figura N° 106: Flujo de control paquete Integration Services (SSIS)



### 5.7.7. Creación del proyecto Analysis Services (SSAS)

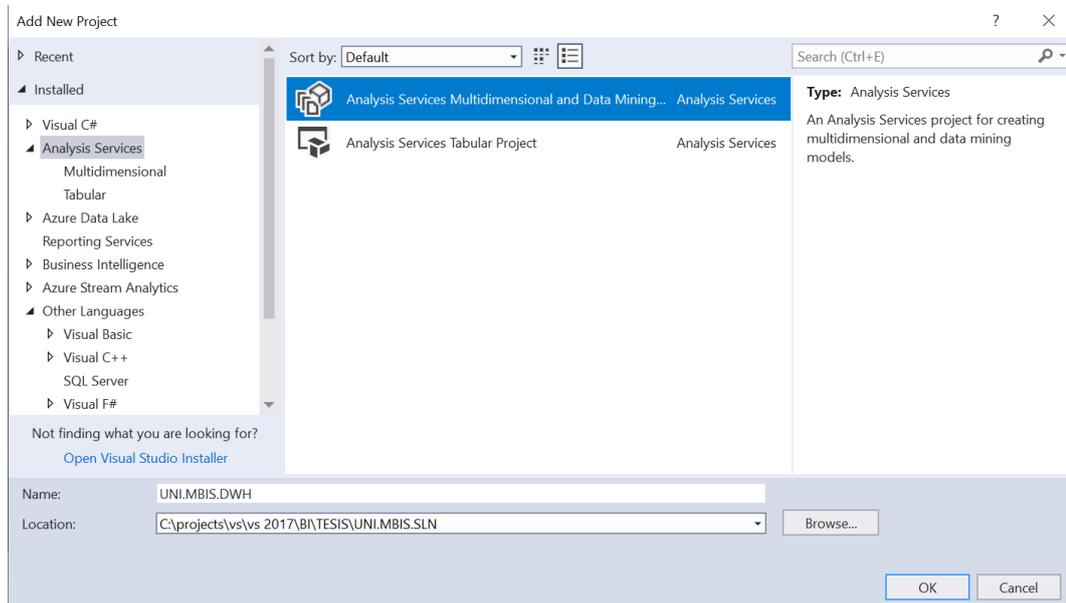
Existen muchas formas de creación de un nuevo proyecto Analysis Services desde la herramienta IDE Visual Studio 2017. En la presente investigación utilizaremos las opciones de menú desplegable. Para iniciar la creación de un proyecto se debe abrir la ventana de dialogo, utilizando la opción Archivo del menú principal y luego seleccionar Crear nuevo proyecto (Figura N° 107).

Figura N° 107: Creación de un proyecto Analysis Services



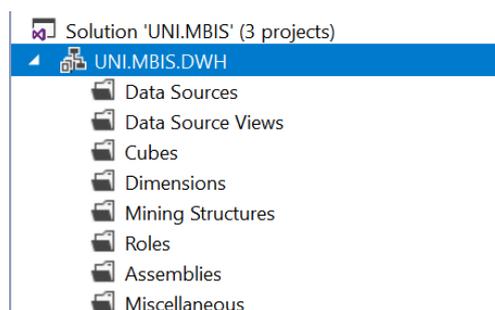
En la ventana de dialogo Crear nuevo proyecto, se puede visualizar que en el panel lateral izquierdo aparecen una lista de categorías de plantillas de proyecto disponibles, las cuales se encuentran organizadas por lenguaje de programación y por tipo de proyecto. Se debe ubicar en la lista, la categoría: Analysis Services, luego seleccionar la plantilla de proyecto: Analysis Services Multidimensional and DataMining, finalmente dar el nombre al proyecto: WTW.MBIS.DWH (Figura N° 108).

Figura N° 108: Selección de plantilla de proyecto Analysis Services



Como se puede apreciar en la Figura N° 109, la herramienta SQL Server Data Tools creó el proyecto UNI.MBIS.DWH de acuerdo con la plantilla de proyecto: Analysis Services, el cual está compuesto de una serie de carpetas de trabajo.

Figura N° 109: Carpetas de plantilla de proyecto Analysis Services

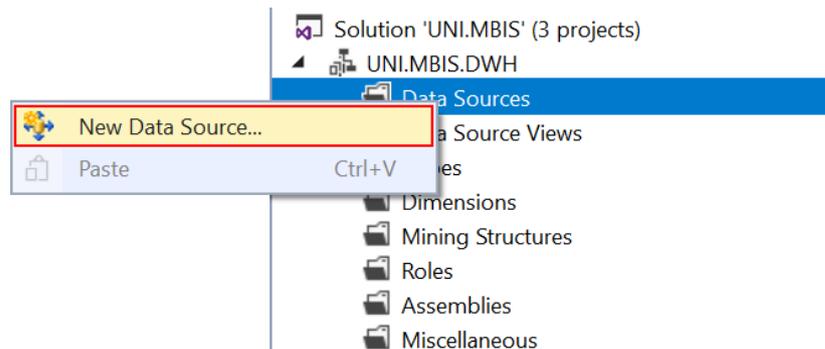


#### 5.7.7.1. Acceso a la fuente de datos

Para tener acceso a la base de datos relacional donde encuentra el almacén de datos, necesitamos crea una conexión a esta base de datos. Desde la

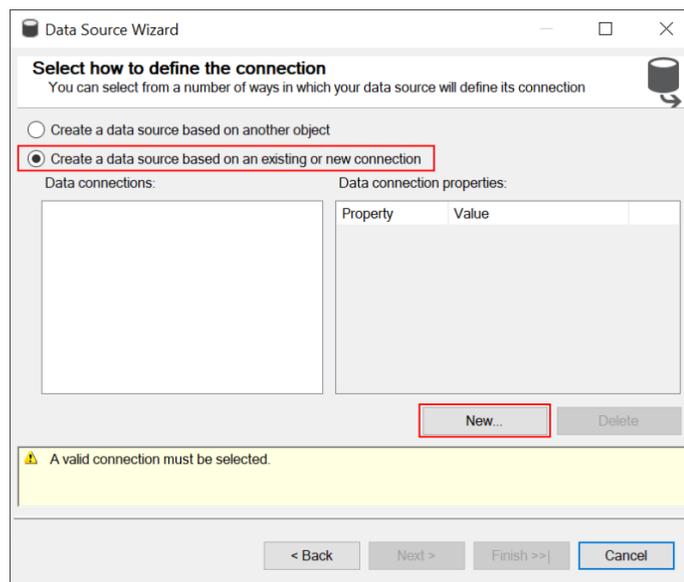
ventana explorador de soluciones dar click derecho sobre la carpeta [Data Sources] luego seleccionar la opción [New Data Source...] (Figura N° 110)

Figura N° 110: Creación de una nueva fuente de datos



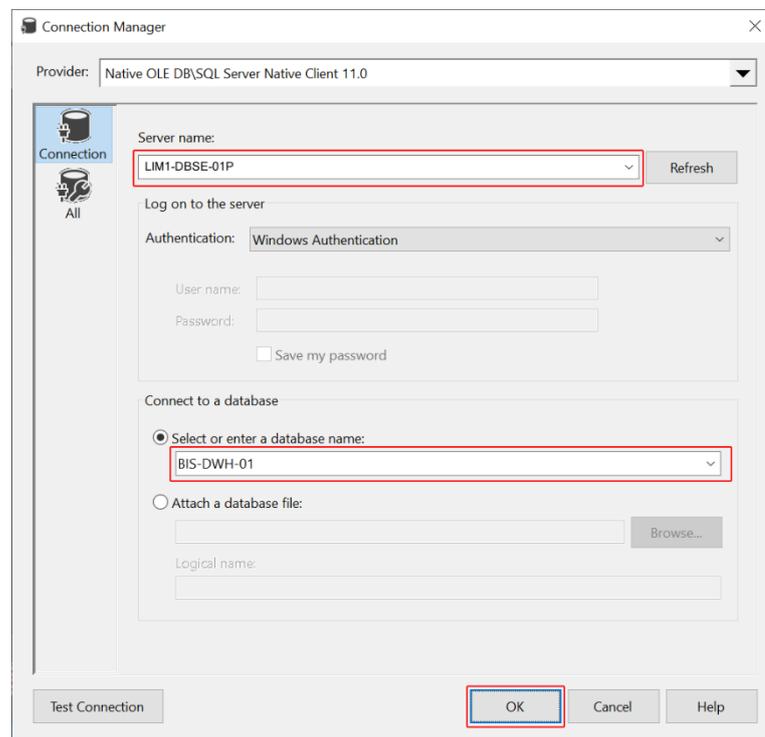
Luego se abrirá la ventana Asistente de Fuente de Datos en el cual nos permitirá definir la cadena de conexión a la base de datos relacional. Seleccionar la opción multiple: [Create a data source based on an existing or new connection] luego dar click sobre el botón [New...] (Figura N° 111)

Figura N° 111: Ventana asistente de fuente de datos



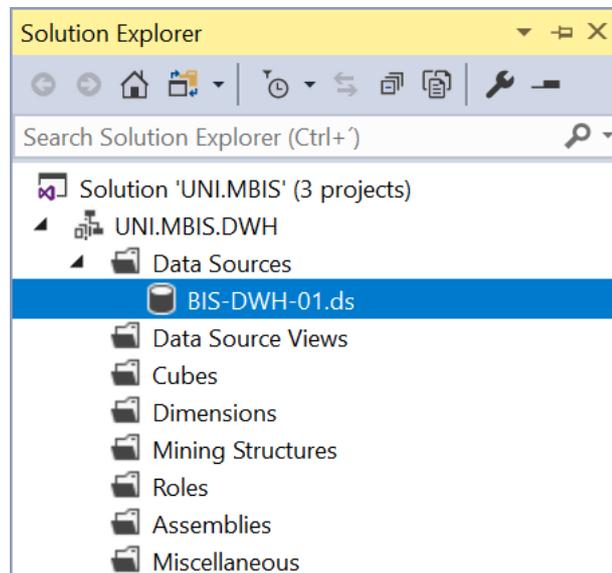
En la ventana de administrador de conexiones se debe ingresar el nombre del servidor de base datos y luego el seleccionar el nombre de la base de datos relacional y finalmente dar click sobre el botón [OK] para confirmar la selección (Figura N° 112).

Figura N° 112: Administrador de conexiones



Finalmente, en la ventana: [Solution Explorer], se podrá visualizar la nueva fuente de datos creada, la cual se denomina: BIS-DWH-01.ds (Figura N° 113)

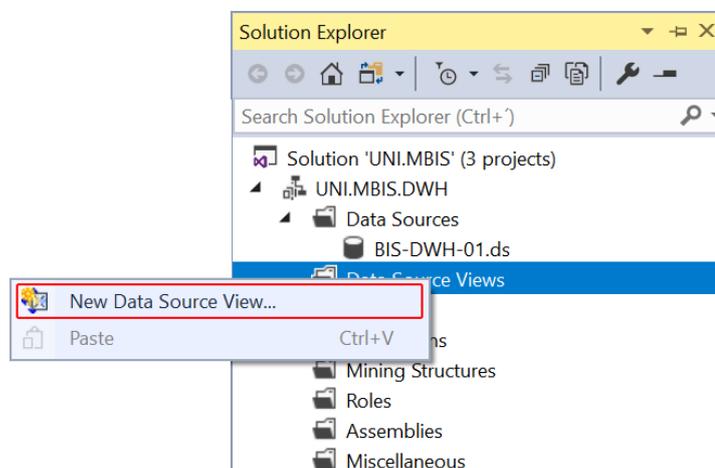
Figura N° 113: Ventana explorador de soluciones



#### 5.7.7.2. Vista de fuente de datos

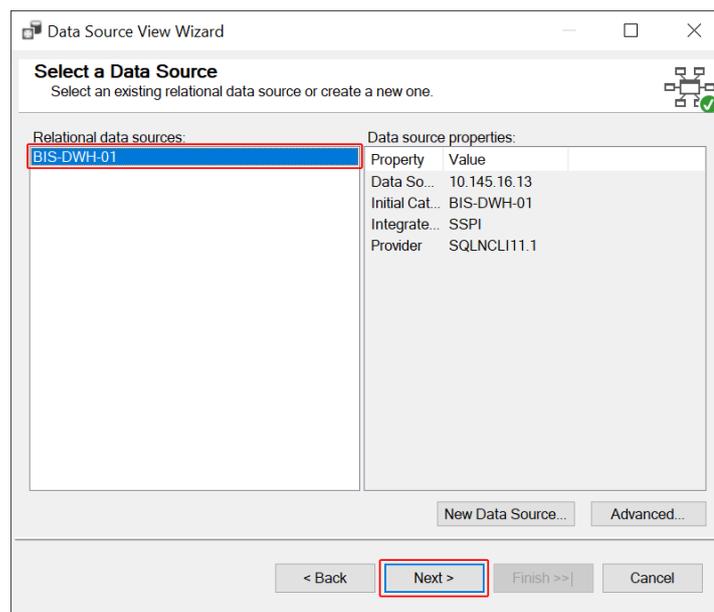
El siguiente paso será la creación de la vista a la fuente de datos. En la ventana: [Solution Explorer] ubicar la carpeta: [Data Source Views] y dar click derecho sobre esta carpeta, inmediatamente se abrirá el menú contextual en el cual se debe seleccionar la opción: [New Data Source View...] (Figura N° 114)

Figura N° 114: Crear nueva vista de fuente de datos



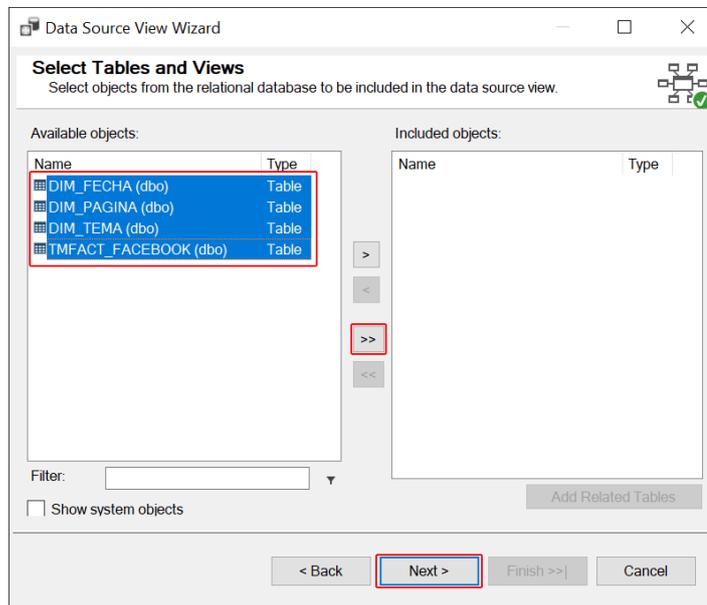
Inmediatamente se abrirá la ventana del Asistente Data Source View en la cual seleccionaremos la conexión a la base datos relacional creada en el paso anterior y luego click sobre el botón [Next>] (Figura N° 115)

Figura N° 115: Ventana vista fuente datos, seleccionar fuente datos



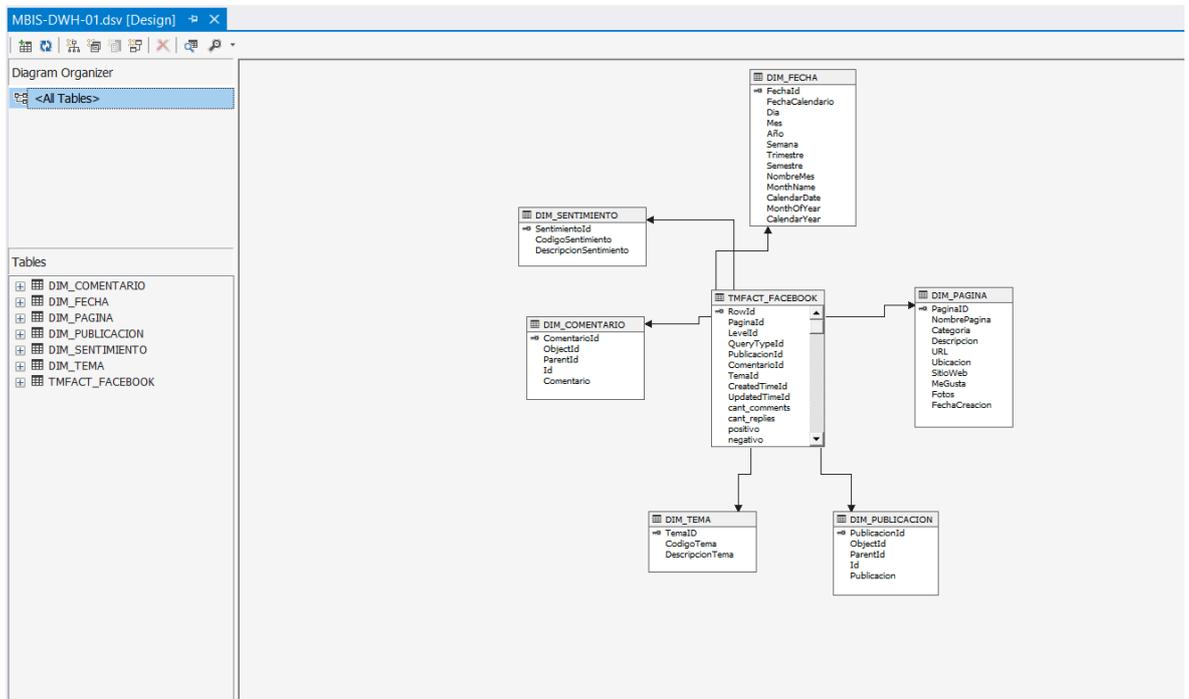
Luego se abrirá la venta del Asistente de Vista de Fuente de Datos, en la cual podremos seleccionar las tablas de dimensiones y tabla de hechos que formarán parte de nuestro modelo dimensional. Para mover las tablas seleccionadas desde el panel lateral izquierdo al panel lateral derecho dar click sobre el botón [>>] y finalmente para confirmar la selección dar click sobre el botón [Next>] (Figura N° 116)

Figura N° 116: Ventana vista fuente datos, seleccionar tablas y vistas



Finalmente, desde la ventana del diseñador de vista de fuente de datos, podemos visualizar el esquema estrella del modelo dimensional, la cual está formada por las tablas de dimensiones, la tabla de hechos y las relaciones entre las tablas de dimensiones (Figura N° 117).

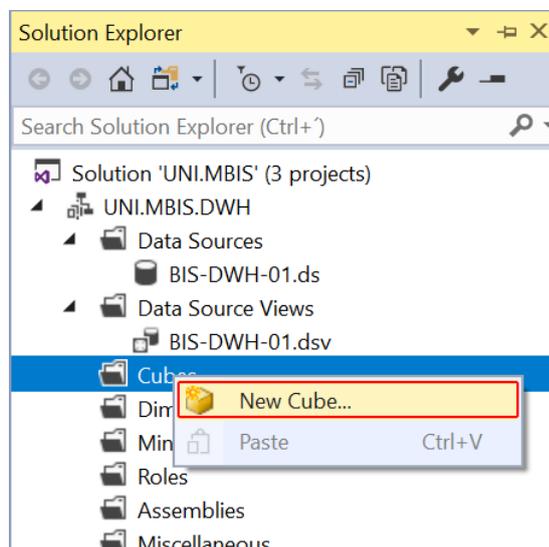
Figura N° 117: Diseñador de vista fuente de datos



### 5.7.7.3. Creación de un nuevo cubo

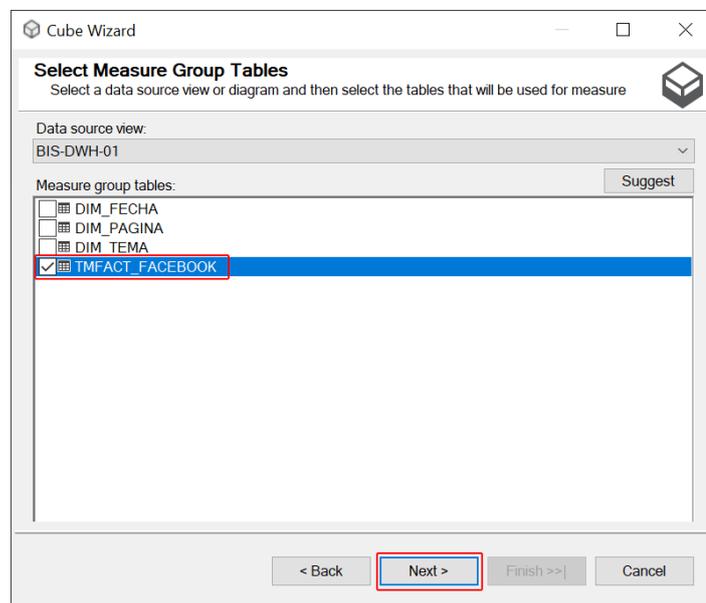
Para iniciar el proceso de creación de un nuevo Cubo dar click derecho sobre la carpeta [Cubes] y luego click sobre la opción: [New Cube...] (Figura N° 118)

Figura N° 118: Creación de un nuevo Cubo



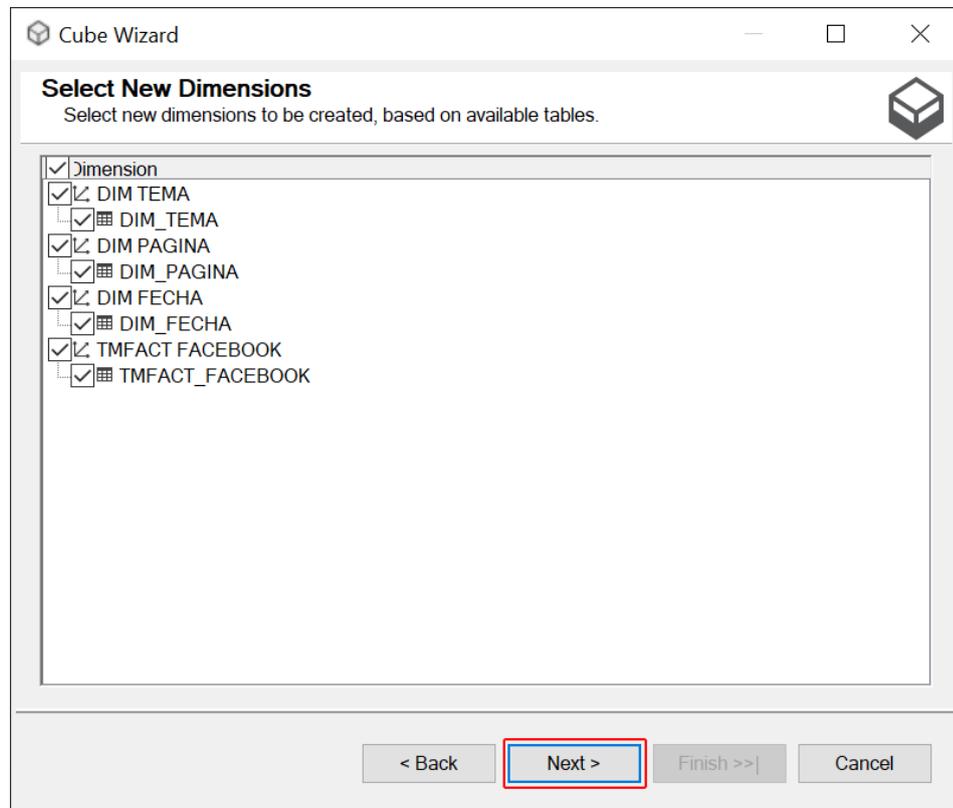
Desde la ventana: Asistente de Cubo, seleccionar la tabla de hechos, que es la tabla central del modelo dimensional, la cual está compuesta de campos clave foránea y de medidas del negocio los cuales se convertirán en medidas de grupo que nos permitirán: contar, acumular, promediar, etc. (Figura N° 119)

Figura N° 119: Ventana asistente de cubo, Seleccionar tablas de medidas de grupo



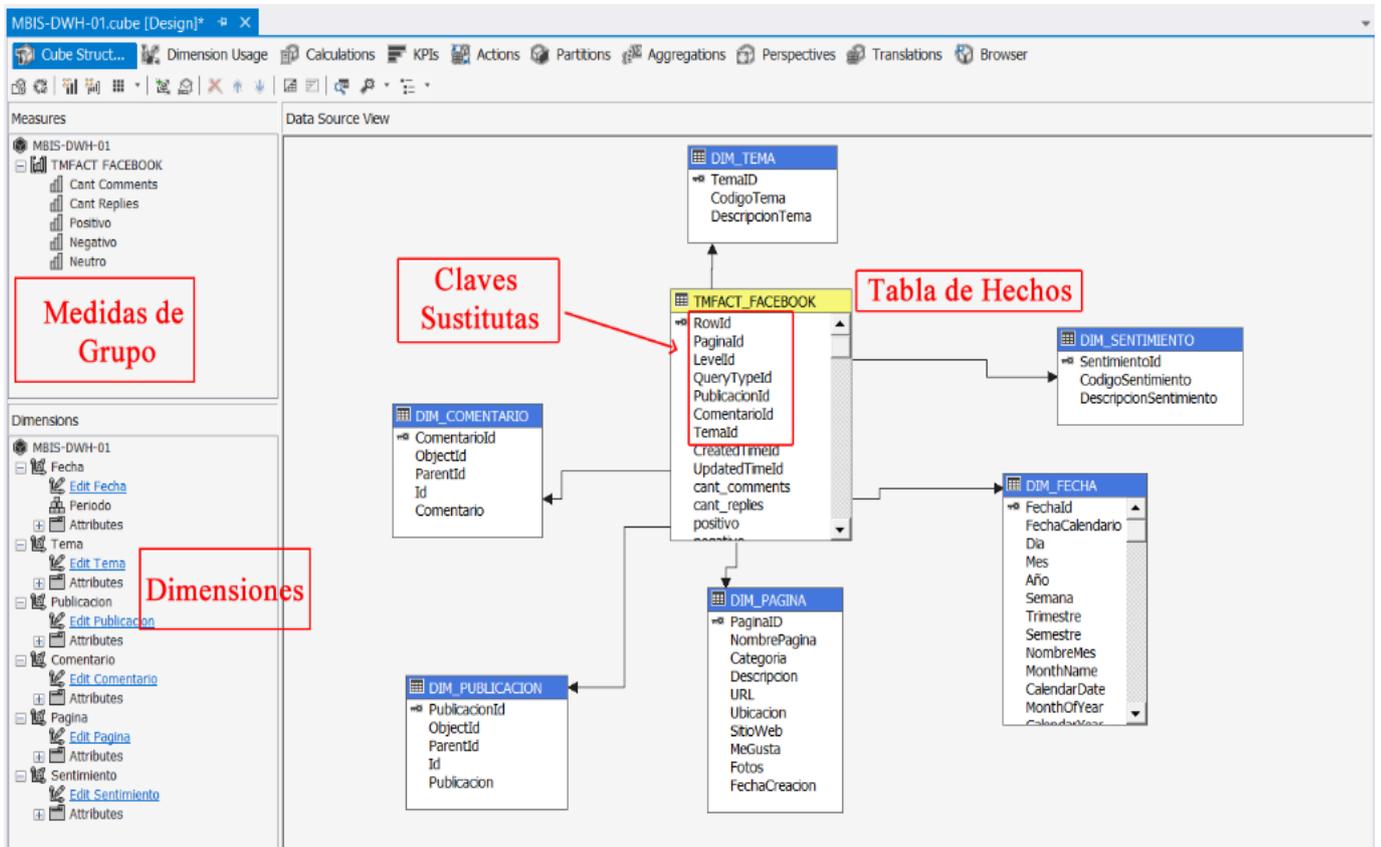
El siguiente paso en el proceso de creación del cubo, es la selección de las dimensiones del cubo, en este caso son: Fecha, Tema y Página. (Figura N° 120)

Figura N° 120: Ventana asistente de cubo, Seleccionar dimensiones



Como se puede apreciar en la Figura N° 121, desde la ventana Diseñador de Cubo podemos visualizar el esquema estrella del cubo, el cual está compuesto de las medidas de grupo, las dimensiones y la tabla de hechos al centro del modelo. La tabla de hechos está relacionada con las tablas de las dimensiones a través de las claves subrogadas.

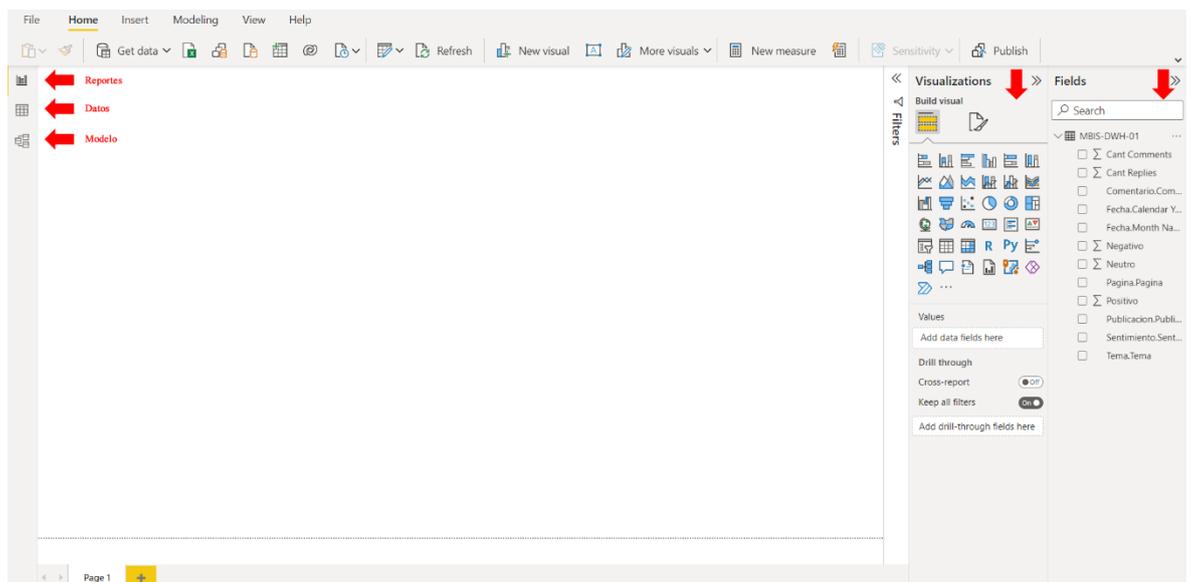
Figura N° 121: Diseñador del cubo



### 5.7.8 Creación del tablero de control en Power BI

Para crear el tablero de control utilizaremos herramienta Microsoft Power BI Desktop, cuyo editor se puede visualizar en la Figura N° 122, en el editor podemos encontrar las diferentes opciones para crear: Reportes, Tabla de Datos y Modelo de Datos en el panel lateral izquierdo. Así mismo se puede visualizar al lado derecho los paneles de Visualización y de Campos de la base datos.

Figura N° 122: Editor de reportes en Power BI



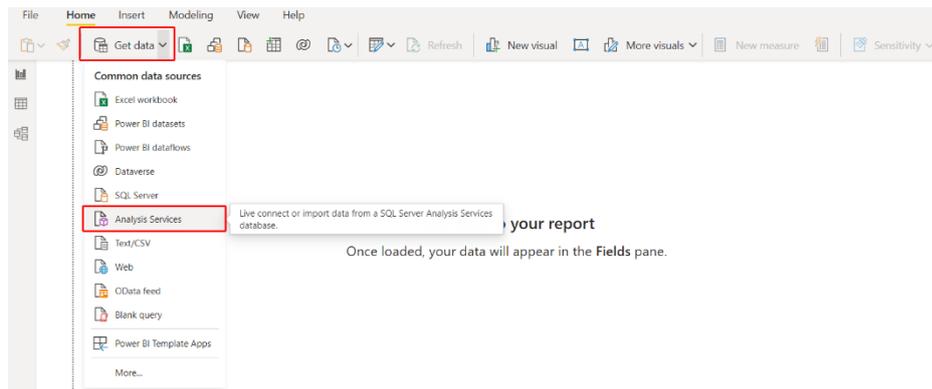
#### 5.7.8.1. Acceso a la fuente de datos

La primera tarea que debe realizar cuando se inicia a crear un tablero de control en Power BI es importar datos de archivos de datos, bases datos relaciones y bases datos dimensionales.

En este caso necesitamos conectarnos a una base datos dimensional (cubo) la cal se encuentra en un servidor OLAP. Para poder tener acceso a la fuente de datos de su elección haga click sobre la opción [Get data].

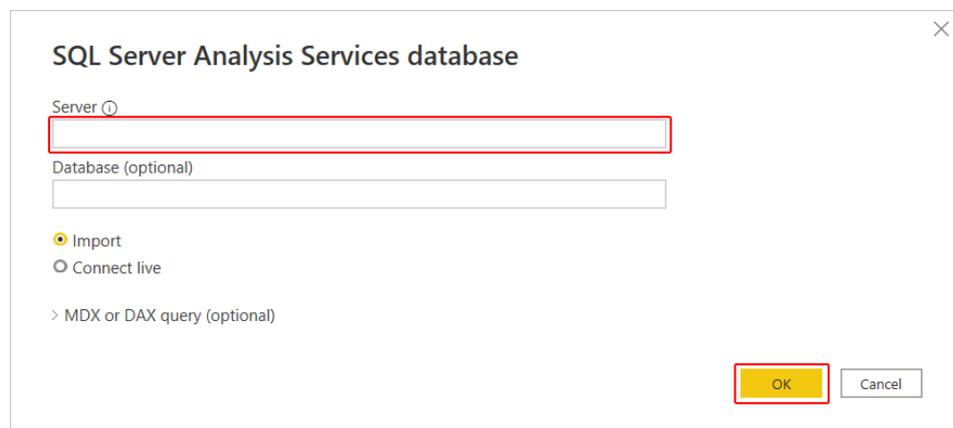
Como se puede apreciar en la Figura N° 123, inmediatamente se despliega un submenú donde se muestran todas las opciones de acceso a datos más comunes, seleccionar la opción: [Analysis Services]

Figura N° 123: Obtener acceso a la fuente de datos



Como se muestra en la Figura N° 124, se abre la ventana: [SQL Server Analysis Services database] en donde podemos visualizar los campos de texto para poder ingresar los parámetros de conexión como: nombre del servidor OLAP y nombre de la base datos.

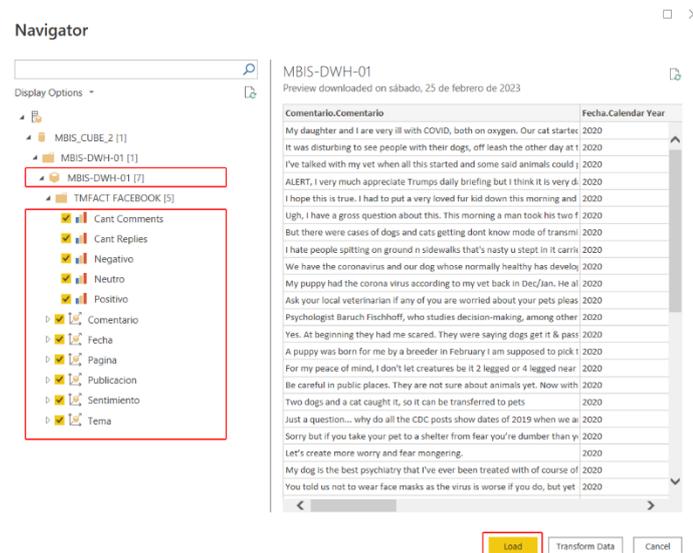
Figura N° 124: Ventana SQL Server Analysis Services Database



Se debe ingresar la IP o nombre del servidor OLAP en donde se encuentra almacenado el cubo y luego dar click sobre el botón [OK]

Como se puede visualizar en la Figura N° 125, luego se abre la ventana de navegación en donde se muestra la estructura que tiene el cubo. Desplegar las carpetas que se encuentran dentro de la carpeta: [MBIS\_CUBE\_2] hasta ubicar la carpeta que representa al cubo: [MBIS-DWH-01].

Figura N° 125: Ventana de navegación del cubo



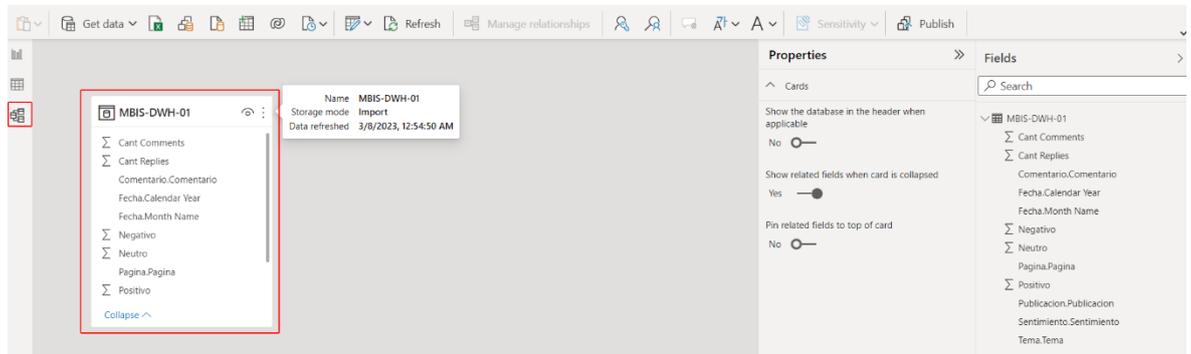
Dentro de esta carpeta se encuentra tabla de hechos: [TMFACT\_FACEBOOK] y las dimensiones. Seleccionar las medidas de grupo y las dimensiones, luego para cargar los datos al modelo de datos, se debe haga click sobre el botón [Load].

### 5.7.8.2. Vista de la tabla de datos

Para poder acceder a las tablas de datos importadas en el paso anterior y verlas en formato tabular, haga click sobre la pestaña [Datos]. Como se



Figura N° 127: Vista del modelo de datos



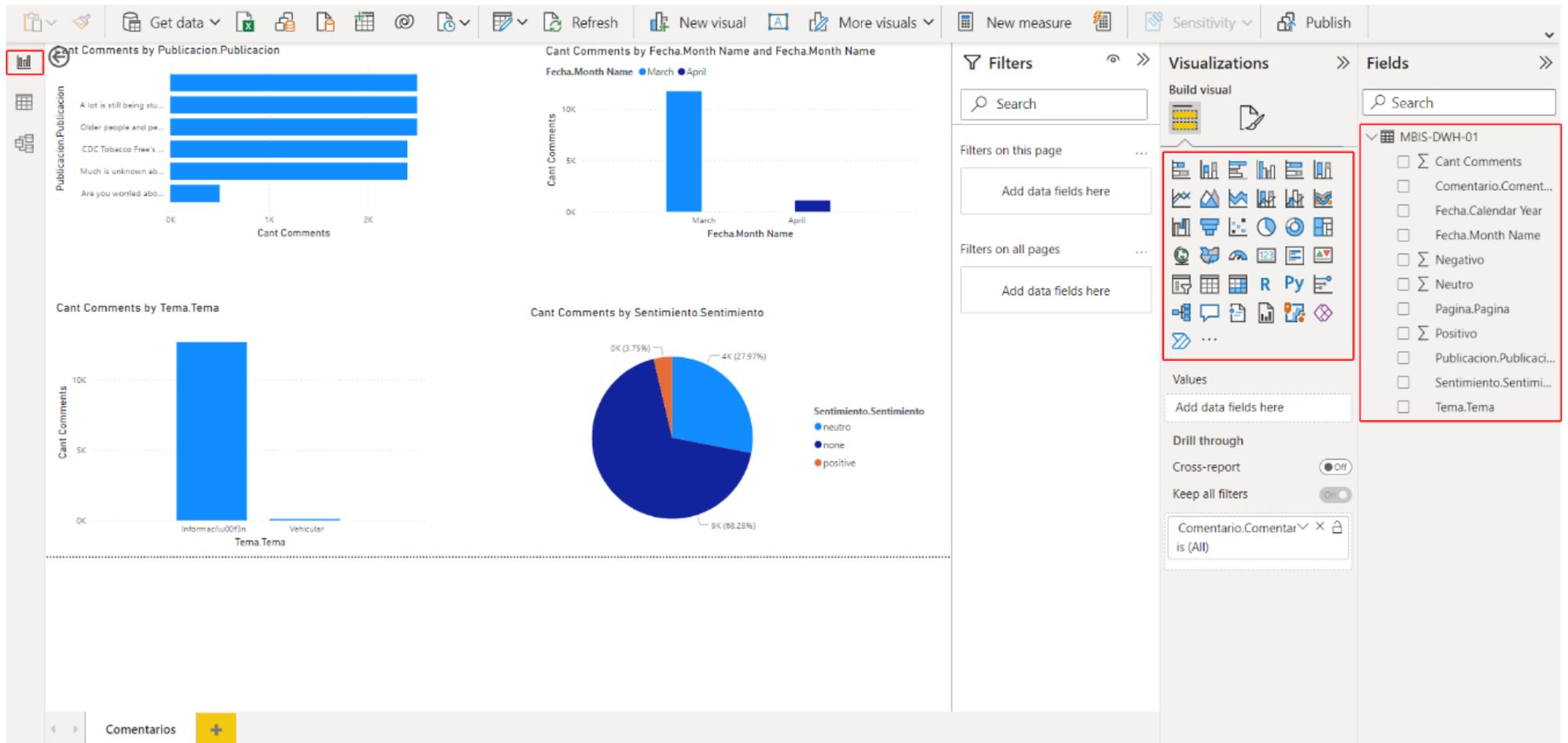
#### 5.7.8.4. Tablero de control

Para poder visualizar el tablero de control se debe dar click sobre la pestaña [Reportes] desde el panel lateral izquierdo. Como se puede visualizar en la Figura N° 128, podemos encontrar los diferentes gráficos como: barras, charts, pye chart, etc.

Los indicadores que podrá encontrar en el tablero de control son:

- Cantidad de Comentarios por Publicación
- Cantidad de Comentarios por Mes
- Cantidad de Comentarios por Temas
- Cantidad de Comentarios por Sentimiento

Figura N° 128: Tablero de control en Power BI



## **CAPÍTULO VI**

### **ANÁLISIS ECONOMICO FINANCIERO**

#### **6.1. INVERSIÓN DEL PROYECTO**

En la implementación del modelo de Inteligencia de Negocios basado en minería de opinión sobre redes sociales, se requiere de la inversión en: recursos humanos (Tabla N° 19), recursos de hardware (Tabla N° 20) y software (Tabla N° 21).

Uno de los costes más importantes del proyecto son las horas/hombre que se requieren para la implementación. El personal técnico que se requiere debe ser calificado y con conocimiento suficiente el manejo de herramientas de desarrollo de: Inteligencia de Negocios, Machine Learning, Construcción de ETL y Construcción de Datamarts.

Tabla N° 19: Costos del proyecto - recursos humanos

N°	Tarea	Responsable	Trabajo (días - hombre)	Horas Hombre	Costo S/.
1	Monitoreo, Control y Seguimiento del Proyecto	Lider del Proyecto	40	320	29,167
2	Diseño Lógico y Físico del Modelo de Datos: Base Datos: Data Stage Base Datos: ODS Base Datos: DWH	Analista OLAP	15	120	5,000
3	Despliegue de Bases Datos Creación de Bases Datos Creación de Tablas Creación de Indices y Estadísticos Afinamiento de Servidores de Base Datos Afinamiento de Servidores de Analysis Services	Administrador Base Datos	10	80	2,917
4	Construcción Paquetes de Extracción de datos: Instalación de Paquetes Facebook.sdk Desarrollo de componente de extracción API	Desarrollador Python	50	400	12,083
5	Construcción Paquetes de Procesamiento de Datos: Programas Rastreadores Modelo de Minería de Texto Modelo de Análisis de Sentimiento	Data Scientist	60	480	14,500
6	Construcción de Paquetes ETL: Parámetros de Carga, Cadenas de Conexión Flujos de Trabajo Limpieza y Estandarización de datos Carga Base Datos Stage, ODS, DWH	Analista ETL	30	240	12,500
7	Construcción del Datamart: Construcción de Dimensiones y Jerarquías Construcción de Tablas de Hechos Construcción de Medidas y KPI's Construcción de Particiones y Agregaciones	Analista ETL	20	160	6,667
8	Construcción de Jobs de Automatización: Programación de Paquetes de Extracción y Carga Programación de Paquetes de Procesamiento	Administrador Base Datos	2	16	583
9	Despliegue de la Solución Servidores de Producción	Analista OLAP	5	40	3,646
10	Generación Reportes de Consulta: Base Datos ODS Datamart	Administrador Base Datos	5	40	1,458
11	Capacitación a Usuarios en el uso de la Herramienta	Capacitador	5	40	1,250
	<b>TOTAL</b>		<b>242</b>	<b>1936</b>	<b>89,771</b>

Fuente: Elaboración Propia (2020)

Tabla N° 20: Costos del proyecto - recurso de hardware

Role Servidor	Sistema Operativo	Recursos			Costo Annual S/.
		GB RAM	CPU	Disco	
Servidor Base Datos	Windows Server 2016 SQL Server 2016 Database Engine SQL Server 2016 Integration Services SQL Server 2016 Reporting Services	32	16	1TB	6,000
Servidor OLAP	Windows Server 2016 SQL Server 2016 Analysis Services	8	4	250GB	4,000
Servidor Archivos	Windows Server 2016	4	4	250GB	2,000
<b>TOTAL</b>					12,000

Fuente: Elaboración Propia (2020)

Tabla N° 21: Costos del proyecto - recurso de software

Herramientas Desarrollo	Costo Licencia Anual S/.
Visual Studio 2017 Professional	5,000
SQL Server Data Tools	0
<b>TOTAL</b>	5,000

Fuente: Elaboración Propia (2020)

## 6.2. RESUMEN DE LOS COSTOS DEL PROYECTO

Los costos del proyecto de implementación del modelo de inteligencia de negocios se encuentran agrupados en recursos Humanos, Hardware y Software, lo cuales detallamos a continuación en la Tabla N° 22.

Tabla N° 22: Resumen de costos del proyecto

Tipo de Recurso	Costo S/.
Recursos Humanos	89,771
Recurso de Hardware	12,000
Recurso de Software	5,000
<b>TOTAL</b>	106,771

Fuente: Elaboración Propia (2020)

### **6.3. RENTABILIDAD DEL PROYECTO**

Para la elaboración de la proyección del Estado de Ganancias y Pérdidas con la información actual se ha considerado el porcentaje de crecimiento de los últimos tres años sobre las ventas y para los gastos se utilizaron los porcentajes de participación que tienen estos conceptos sobre las ventas.

Para la elaboración de la proyección del Estados de Resultados proyectados se ha considerado un porcentaje de crecimiento de 8% en ventas, mientras para los gastos se han utilizado los porcentajes de participación de estos conceptos sobre las ventas.

Para calcular el valor presente de los flujos se ha utilizado la diferencia de la utilidad neta de ambos estados de resultados, luego se actualizaron ambos flujos y se trajeron a valor presente.

Para medir la rentabilidad del proyecto utilizaremos los siguientes indicadores financieros los cuales nos indicaran la factibilidad en la implementación del proyecto:

- Valor Actual Neto (VAN)
- Tasa Interna de Retorno (TIR)

Como se puede apreciar en la Tabla N° 23, el resultado obtenido es una VAN mayor a cero (S/. 61M) y una TIR de 37% por lo tanto podemos concluir que el proyecto es rentable y es factible su implementación en la empresa.

Tabla N° 23: Cálculo del valor presente de los flujos (en miles de nuevos soles)

Concepto	Año Cero	2020	2021	2022	2023
Utilidad Neta Actual		1,865	2,014	2,175	2,349
Utilidad Neta Proyectada		1,902	2,065	2,230	2,408
<b>Diferencia</b>	-92	37	50	54	59

VAN	61
TIR	37%

Fuente: Elaboración Propia (2020)

## **CAPÍTULO VII**

### **ANÁLISIS Y DISCUSIÓN DE RESULTADOS**

#### **7.1. RESULTADOS DE LA INVESTIGACIÓN**

En este capítulo presentamos los principales resultados encontrados durante la investigación en los cuales se muestra la importancia que tiene la información y de manera se ha convertido en uno de los activos más importantes que tienen las organizaciones y muchas veces no son conocidas por las organizaciones por lo tanto se hace imprescindible por extraerlas organizarlas y tomar decisiones basados en datos y de manera oportuna.

##### **7.1.1. Extracción de comentarios**

En los últimos años en el mundo la generación de datos no estructurados (Big Data) se ha incrementado considerablemente, actualmente representan el 80% en comparación a su contraparte los datos estructurados que representan solo el 20%.

Los datos no estructurados se encuentran dispersos en entornos externos a las organizaciones como son las redes sociales, significando un alto riesgo para su reputación si es que no son extraídas y procesadas a tiempo.

Existen tres motivaciones para que los usuarios estén dispuestos compartir contenido en redes sociales:

- **Social:** Enviar señales a otros sobre la experiencia, la singularidad o el estatus social.
- **Emocional:** Compartir sentimientos positivos o negativos sobre un producto o servicio.
- **Funcional:** La Tendencia de obtener y proporcionar información a otros

En el mundo existen 2,796 billones de usuarios se conectan mensualmente a Facebook lo cual representa un incremento de +11% en comparación al Q4 2019. Adicionalmente en Facebook, cada 60 segundos se publican 510,000 comentarios, se actualizan 293,000 estados y se suben 136,000 fotos.

En el Perú existen 24 millones de usuarios activos en redes sociales lo cual representan un 73% de la población, 23 millones de usuarios utilizan un teléfono inteligente para acceder a internet lo cual representa un 70% de la población. Adicionalmente existen 39.08 millones de dispositivos móviles lo cual representa un 119% en función a la población.

En el Perú la inversión en publicidad digital se incrementó un 6% en el 2020 en donde los formatos de publicidad que más crecieron fueron: video creció un +6%, redes sociales +7% y display cayó -25%

### **7.1.2. Clasificación de texto**

La clasificación o categorización de texto es el proceso de asignar un documento en una categoría o clase de acuerdo con las características del documento de entrada.

La clasificación de texto se ha utilizado para una serie de aplicaciones las cuales no se usarán en la presente investigación como la categorización del habla, categorización de documentos multimedia.

Para las tareas de clasificación de texto se considera dos tipos de enfoques: basado en reglas y basado en Machine Learning, en la presente investigación nos hemos basado en el enfoque basado en Machine Learning.

Los datos extraídos desde las redes sociales serán clasificados en una categoría utilizando un algoritmo de clasificación, en la presente investigación los algoritmos de clasificación de texto que será evaluados son: Naive Bayes, Decision Tree y Support Vector Machine (SVM).

Como se puede visualizar en la Figura N° 75, después de someter a evaluación los algoritmos de clasificación y después de comparar los resultados, el algoritmo que tiene el mejor valor de precisión es SVM

(Support Vector Machine) con 78.99%, en segundo lugar, se ubica el algoritmo Naive Bayes con una precisión del 74,78% y finalmente el algoritmo Decision Tree que alcanza una precisión del 57,66%.

Después de evaluar el funcionamiento de los dos herramientas disponibles para el proceso de extracción de comentarios: API Graph y Facepager, se ha tomado la decisión de utilizar la herramienta Facepager esto debido a que la curva de aprendizaje es menor y que está continuamente en actualización de la herramienta, en comparación con la API Graph que es una herramienta proveída por Facebook para el uso de los desarrolladores en sus proyectos pero que su uso se hace más complejo debido a que Facebook cambia contantemente sus políticas de seguridad sobre el acceso a los datos públicos.

### **7.1.3. Análisis de sentimiento**

El análisis de sentimiento es una de las tareas dentro de la clasificación de texto que consiste en usar lenguaje natural (también es conocido como NLP) y Machine Learning para interpretar y clasificar emociones expresadas en texto.

El análisis de sentimiento se realiza en tres niveles diferentes:

- Nivel de Documento, clasifica si la opinión del documento es: positiva, negativa o neutra.
- Nivel de Oración, determina si la oración expresa alguna opinión positiva, negativa o neutra

- Nivel de Aspecto, se enfoca en todas las expresiones de sentimientos presentes en un documento dado

El análisis de sentimiento también se puede aplicar a:

- Análisis de sentimiento en los textos
- Análisis de sentimiento en el habla
- Análisis de sentimiento visual

Para poder conocer, entender y clasificar emociones expresadas a través de comentarios en redes sociales, se necesita transformar los comentarios en información objetiva para la toma de decisiones.

Para transformar comentarios en sentimiento: positivo, negativo o neutro se ha desarrollado el modelo de Análisis de Sentimiento basado en Machine Learning, utilizando las librerías: Tensorflow y Keras.

Tensorflow es una de las librerías de Deep Learning que utiliza una red neuronal detrás, además es una de las librerías más utilizadas para realizar tareas de análisis de sentimiento, cuya implementación y curva de aprendizaje es bastante corta.

Cómo se puede apreciar en la Figura N° 135, El modelo de Análisis de Sentimiento tiene una precisión 76% para predecir si un comentario tiene un sentimiento: positivo, negativo o neutro.

Si bien es cierto el Modelo de Análisis de Sentimiento no brinda un 100% de precisión, pero es una herramienta útil para medir el nivel de satisfacción de

los clientes ante un producto o servicio y además permite detectar campañas desprestigio a tiempo en redes sociales, los cuales deben tratarse de forma inmediata para evitar que la imagen de empresa se vea perjudicada.

Como lo sucedido con la empresa BMW en julio del 2022, tuvo un alza de sus menciones en Twitter, pero las interacciones que expresaban los usuarios eran mayormente negativas por la aparente decisión de la empresa de vender servicios de suscripción para algunas funciones internas de los carros. El tuit que originó la debacle aglomeró 27,000 retuits y 217,000 me gusta, tal como se puede apreciar en la Figura N° 129, a tal punto que la empresa BMW USA se vio obligada a aclarar los planes de suscripción a través de otro tuit Figura N° 130 y un comunicado oficial.

Figura N° 129: Tuit originó la debacle



Fuente: Twitter

Figura N° 130: Tuit aclaración BMW USA



Fuente: Twitter

#### **7.1.4. Tablero de control**

Las plataformas de Inteligencia de Negocios y Analítica permiten a los usuarios con menor capacidad técnica como: empresarios y analistas de negocios, a modelar, analizar, explorar, compartir, administrar datos, colaborar y compartir sus hallazgos.

Como se muestra en la Figura N° 131, La consultora internacional Gartner ha reconocido a Microsoft en el año 2022 como líder de plataformas de Inteligencia de Negocios y Analítica en su Cuadrante Mágico.

El fabricante Microsoft logra este reconocimiento por cuarto año consecutivo, y se vuelve a ubicar en lo más alto del cuadrante en cuanto a Integridad de la Visión y Capacidad de Ejecución.

Podemos visualizar en el gráfico que las distintas soluciones de Inteligencia de Negocios se encuentran divididos en 4 categorías:

- Retadores
- Jugadores de Nicho
- Visionarios
- Lideres

Estas categorías, a su vez, se ubican dentro del cuadrante en relación con los dos ejes principales:

- Eje X (Integridad de la Visión): habilidad de los proveedores para leer correctamente el contexto del mercado en el que se encuentran.
- Eje Y (Capacidad de Ejecución): capacidad de estos proveedores para ejecutar con la mayor eficiencia posible esta lectura del mercado.

Figura N° 131: Cuadrante Mágico Gartner para plataformas de Inteligencia de Negocios y Analítica



Fuente: Gartner (March 2022)

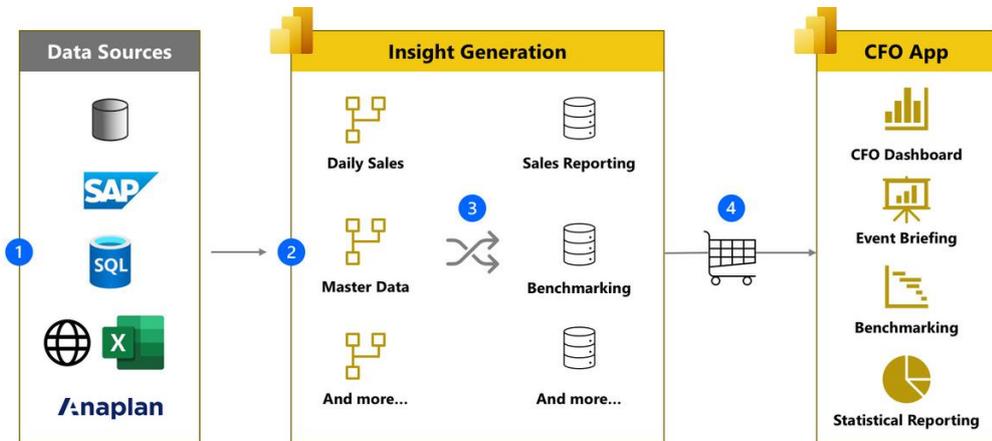
Podemos visualizar que las herramientas líderes de Inteligencia de Negocios y Analítica que se ubican en la parte superior derecha del cuadrante serán aquellas que mejor cumplan estos dos requisitos mencionados.

Una de las herramientas que cumple estos dos requisitos mencionados es Microsoft Power BI la cual se encuentra integrados con Office 365, Microsoft Teams, Excel y SharePoint. La visión que tiene Microsoft que Power BI se convierta en el centro de datos y análisis de una empresa.

En el tablero de control que se puede visualizar en la en la Figura N° 128, podemos encontrar los reportes acumulados por cantidad de comentarios por publicación, la tendencia que muestran de los comentarios por meses, la cantidad de comentarios por temas y el porcentaje de sentimiento (positivo, negativo o neutro) de que tienen los comentarios respecto a una publicación.

Algunos casos de éxito que se publican en la página (Microsoft, 2022) en el uso de la herramienta Power BI se puede encontrar a la empresa alemana Bayer. Bayer es una de las empresas de ciencias de la vida más grandes del mundo (ciencia de cultivo, productos farmacéuticos y salud del consumidor), está presente en 87 países y cuenta con 103,824 empleados al rededor del mundo.

Figura N° 132: Arquitectura de flujo de datos de Bayer



- 1 Roughly 50 data sources (internal and external) imported.
- 2 Roughly 20 Power BI dataflows group data into reusable building blocks and enable automation.
- 3 Power Query extracts data from building blocks for targeted insights.
- 4 Microsoft and AppSource visuals display generated insights in Power BI reports.

Como se puede apreciar en la Figura N° 132, Bayer crea la aplicación CFO para una ventanilla única mediante la gestión de datos relevantes a través de la herramienta Power BI, todos los descubrimientos de los datos y entendimiento del negocio se publican en la aplicación CFO, que actualmente incluye cuatro paquetes de informes:

1. El tablero CFO proporciona una descripción holística sobre la dirección KPI relevante
2. Paquetes informativos de eventos que resumen un evento de planificación
3. Informe de evaluación comparativa que compara a Bayer con sus compañeros y el mercado
4. Reportes estadísticos

Uno de los impactos más significativos que ha experimentado Bayer es el ahorro de tiempo. Los datos se cargan instantáneamente mensualmente, diariamente o incluso varias veces al día. Debido a que Bayer automatizó este proceso de carga, esencialmente no hay demoras en el data warehouse.

#### **7.1.5. Indicadores del tablero de control**

Para obtener los indicadores que se visualizarán en el tablero de control, se ha construido scripts utilizando el lenguaje de consulta MDX (Multidimensional Expressions), el cual es empleado para manipular y recuperar datos desde grandes bases de datos dimensionales en Analysis Services.

Una base de datos dimensional en Analysis Services está compuesta de: medidas, dimensiones y atributos de las dimensiones. Las medidas son los datos numéricos que podemos encontrar en las celdas, son valores agregados a través de funciones como: suma, contar, porcentaje, promedio, máximo, mínimo. Las dimensiones son información cualitativos, son las que dan sentido a los datos de las medidas.

La sentencia MDX SELECT está compuesta de la cláusula SELECT y la cláusula FROM. En la cláusula SELECT podemos definir las dimensiones y medidas, además de los ejes (filas y columnas) en los que se retornarán los registros que cumplan la condición. En la cláusula FROM podemos definir el

nombre de la base de datos dimensional sobre la que se ejecutará la consulta.

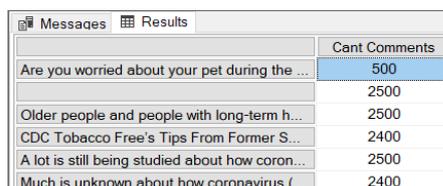
#### 7.1.5.1. Cantidad de comentarios por publicación

Este indicador cuenta la cantidad de comentarios que los usuarios han realizado a una publicación. Este indicador nos permite medir el grado de reacción que ha generado en los usuarios una publicación determinada.

Como se puede visualizar en la Figura N° 133, en el eje de las columnas (ON COLUMNS), podemos encontrar la medida: [Measures].[Cant Comments] y en el eje de las filas (ON ROWS) podemos encontrar la dimensión: [Publicacion].[Publicacion].[Publicacion].

Figura N° 133: Script MDX - Cantidad de comentarios por publicación

```
-- 1.- [Cantidad de Comentarios por Publicacion]
SELECT NON EMPTY { [Measures].[Cant Comments] } ON COLUMNS,
NON EMPTY { ([Publicacion].[Publicacion].[Publicacion].ALLMEMBERS ) } ON ROWS
FROM [MBIS-DWH-01]
```



	Cant Comments
Are you worried about your pet during the ...	500
	2500
Older people and people with long-term h...	2500
CDC Tobacco Free's Tips From Former S...	2400
A lot is still being studied about how coron...	2500
Much is unknown about how coronavirus (...)	2400

Al ejecutar el script sobre la base de datos dimensional [MBIS-DWH-01], retorna un conjunto de datos que nos muestra la cantidad de comentarios en el eje de las columnas y en el eje de las filas las publicaciones.

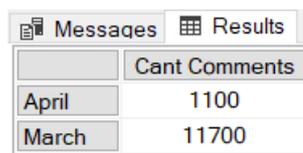
### 7.1.5.2. Cantidad de comentarios por mes

Este indicador cuenta la cantidad de comentarios que los usuarios han realizado a las publicaciones en un determinado mes del año. Este indicador nos permite medir la tendencia que tienen los usuarios en comentar a las publicaciones a través del tiempo.

Como se puede visualizar en la Figura N° 134, en el eje de las columnas (ON COLUMNS) podemos encontrar la medida: [Measures].[Cant Comments] y en el eje de las filas (ON ROWS) podemos encontrar la dimensión: [Fecha].[Periodo].[Month Name].

Figura N° 134: Script MDX - Cantidad de comentarios por mes

```
-- 2.- [Cantidad de Comentarios por Mes]
SELECT NON EMPTY { [Measures].[Cant Comments] } ON COLUMNS,
NON EMPTY { ([Fecha].[Periodo].[Month Name].ALLMEMBERS ) } ON ROWS
FROM [MBIS-DWH-01]
```



	Cant Comments
April	1100
March	11700

Al ejecutar el script sobre la base de datos dimensional [MBIS-DWH-01], retorna un conjunto de datos que nos muestra la cantidad de comentarios en el eje de las columnas y en el eje de las filas, los meses a los que pertenecen.

### 7.1.5.3. Cantidad de comentarios por tema

Este indicador cuenta la cantidad de comentarios que los usuarios han realizado a las publicaciones de un determinado tema. Este indicador nos permite conocer las preferencias que tiene acerca de un tema determinado.

Como se puede visualizar en la Figura N° 135, en el eje de las columnas (ON COLUMNS), podemos encontrar la medida: [Measures].[Cant Comments] y en el eje de las filas (ON ROWS) podemos encontrar la dimensión: [Tema].[Tema].[Tema]

Figura N° 135: Script MDX - Cantidad de comentarios por tema

```
-- 3.- [Cantidad de Comentarios por Tema]
SELECT NON EMPTY { [Measures].[Cant Comments] } ON COLUMNS,
NON EMPTY { ([Tema].[Tema].[Tema].ALLMEMBERS ) } ON ROWS
FROM [MBIS-DWH-01]
```

Messages	Results
	Cant Comments
Información	12700
Vehicular	100

Al ejecutar el script sobre la base de datos dimensional [MBIS-DWH-01], retorna un conjunto de datos que nos muestra la cantidad de comentarios en el eje de las columnas y en el eje de las filas los temas.

#### 7.1.5.4. Cantidad de comentarios por sentimiento

Este indicador calcula el porcentaje de sentimiento positivo, negativo o neutro que han tenido los comentarios en función del total de comentarios. Este indicador nos muestra el porcentaje de participación que tienen un determinado sentimiento en las reacciones de los usuarios acerca de una determinada publicación.

Como se puede visualizar en la Figura N° 136, en el eje de las columnas ON COLUMNS, podemos encontrar las medidas: [Measures].[Cant Comments], y [Measures].[Percentage]. Además, en el eje de las filas (ON ROWS) podemos encontrar la dimensión: [Sentimiento].[Sentimiento].[Sentimiento].

Figura N° 136: Script MDX - Cantidad de comentarios por sentimiento

```
-- 4.- [Cantidad de Comentarios por Sentimiento]
WITH
  MEMBER [Measures].[Percentage] AS
    [Measures].[Cant Comments]/
    ([Sentimiento].[Sentimiento].currentmember.parent,[Measures].[Cant Comments]),
  FORMAT_STRING = '0.0000'
SELECT
  NON EMPTY {[Measures].[Cant Comments],[Measures].[Percentage] } ON COLUMNS,
  NON EMPTY {[Sentimiento].[Sentimiento].[Sentimiento].ALLMEMBERS ) } ON ROWS
FROM [MBIS-DWH-01]
```

	Cant Comments	Percentage
neutro	3580	0.2797
none	8740	0.6828
positive	480	0.0375

Al ejecutar el script sobre la base de datos dimensional [MBIS-DWH-01], retorna un conjunto de datos que nos muestra la cantidad de comentarios y su porcentaje sobre el total de comentarios en el eje de las columnas. En el eje de las filas podemos encontrar el sentimiento.

### **7.1.6. Ventajas**

La empresa en estudio no cuenta con un sistema para analizar las opiniones y emociones expresadas en comentarios en redes sociales que le permita tomar decisiones en forma oportuna y basada en hechos para identificar y corregir problemas en la entrega de servicio en forma oportuna.

La implementación de un modelo de Inteligencia de Negocios basado en información extraída desde redes sociales Facebook, generará las siguientes ventajas en la empresa en estudio:

- **Eliminación de Problemas de Bloqueo y Baja de la Performance en Bases Datos Transaccionales.**

La generación de reportes en la empresa en estudio se realiza de manera aislada y dispersa a través de los analistas de negocios de las gerencias de comercial y de operaciones, el proceso de extracción se realiza a ejecutando de consultas directas a las bases de datos transaccionales, esto muchas veces ocasiona problemas de bloqueo y reducción de la performance en la base de datos debido a la naturaleza de las bases de transaccionales.

- **Optimización del uso de Horas/Hombre**

Para generar reportes de formar manuales para la toma decisiones, los analistas de negocio necesitan invertir horas de trabajo para extraer, organizar, procesar y categorizar grandes cantidades de datos de diferentes fuentes. Esto involucra la inversión de horas de trabajo de profesionales cuya función principal es la de analizar la información y generar reportes para las diferentes gerencias de la organización.

Esto ocasiona desperdicios de tiempo de los analistas de negocios en la generación de reportes y reducción de la eficiencia de su trabajo debido a que utilizan su tiempo en desarrollar tareas manuales.

- **Reducción de Errores**

Debido a la naturaleza del ser humano que está proclive a cometer errores si es que realiza tareas repetitivas por un largo periodo de tiempo, durante las tareas de extracción y procesamiento el analista de negocio está expuesto a cometer errores debido a que se realizan de manera manual. Corregir errores genera desperdicios de tiempo en las organizaciones.

- **Incrementar el poder la negociación con Aseguradoras**

El conocimiento es poder, tener información confiable, oportuna y basada en hechos ayudará a los ejecutivos comerciales, especialista de seguros, ejecutivos de operaciones y ejecutivos de servicio al cliente a tomar mejores decisiones. Tener la información directa grupos de usuarios de seguros, incrementará el poder de negociación que tiene la empresa en estudio con

las compañías aseguradoras para lograr una mejora en la entrega del servicio lo cual mejorará el nivel de servicio al cliente.

## 7.2. HERRAMIENTAS DE INVESTIGACIÓN

Como herramienta de investigación se ha utilizado la encuesta tomando como muestra a 20 empleados de la empresa los cuales serán los usuarios directos del modelo de inteligencia de negocios a implementar.

Como se puede visualizar en la Tabla N° 24, durante el diseño de las encuestas se la utilizado una escala de Likert de tipo frecuencia, la cual cuenta con 5 cinco opciones. Se considera que 4 y 5 representan los valores más altos de las opciones, lo cual significa que los usuarios demuestran mucho más interés en utilizar el modelo para sus procesos de negocios.

Así mismo se considera que los valores 2 y 1 son los valores más bajo de la encuesta, lo cual demuestra que los usuarios no tienen interés en utilizar el modelo de inteligencia de negocios propuesta.

Tabla N° 24: Escala likert de tipo frecuencia

Frecuencia	Valor
Nunca	1
Raramente	2
Ocasionalmente	3
Frecuentemente	4
Muy Frecuentemente	5

Fuente: Elaboración Propia (2020)

### **7.3. PRUEBA DE HIPÓTESIS**

Para contrastar los resultados de la presente investigación se ha planteado las siguientes hipótesis, las cuales serán evaluadas usando la prueba de bondad de ajuste (Anderson, Sweeney, & Williams, 2008).

#### **7.3.1. Prueba de hipótesis general**

Un Modelo de Inteligencia de Negocios basado en minería de opinión sobre redes sociales, contribuye a la mejora del servicio al cliente, en una empresa corredora de seguros en el Perú.

##### **7.3.1.1. Hipótesis nula ( $H_0$ )**

Un Modelo de Inteligencia de Negocios basado en minería de opinión sobre redes sociales, contribuye a la mejora del servicio al cliente, en una empresa corredora de seguros en el Perú.

##### **7.3.1.2. Hipótesis alternativa ( $H_a$ )**

Un Modelo de Inteligencia de Negocios basado en minería de opinión sobre redes sociales, no contribuye a la mejora del servicio al cliente, en una empresa corredora de seguros en el Perú.

##### **7.3.1.3. Encuesta realizada**

Para poder probar la hipótesis general se necesita desarrollar la primera encuesta la cual debe ser llenada por los usuarios del Modelo de Inteligencia de Negocios basado en minería de opinión, la cual podemos encontrar en la Tabla N° 25:

Tabla N° 25: Encuesta 01 – hipótesis general

Pregunta	5	4	3	2	1	Total
La información obtenida del Modelo de Inteligencia de Negocios basado en minería de opinión sobre redes sociales, le ha permitido tomar decisiones acertadas y oportunas?	13	4	2	1	0	20
La información obtenida del Modelo de Inteligencia de Negocios basado en minería de opinión sobre redes sociales, le ha permitido identificar el comportamiento y las opiniones que expresan los usuarios en redes sociales?	15	2	2	1	0	20
La información obtenida del Modelo de Inteligencia de Negocios basado en minería de opinión sobre redes sociales le ha permitido identificar problemas en forma proactiva?	15	2	2	1	0	20
La información obtenida del Modelo de Inteligencia de Negocios le ha permitido mejorar del servicio al cliente?	14	3	2	1	0	20

Fuente: Elaboración Propia (2020)

#### 7.3.1.4. Matriz de frecuencia observada (FO)

A continuación, podemos encontrar la matriz de frecuencia observada (Fo) la cual se puede visualizar en la Tabla N° 26.

Tabla N° 26: Matriz de frecuencia observada (Fo)

5	4	3	2	1	Total
13	4	2	1	0	20
15	2	2	1	0	20
15	2	2	1	0	20
14	3	2	1	0	20
57	11	8	4	0	80

Fuente: Elaboración Propia (2020)

### 7.3.1.5. Matriz de frecuencia esperada (FE)

Para poder obtener la matriz de frecuencia esperada utilizaremos la Ecuación 2

**Fórmula:**

$$E_i = \frac{(n_{i.} * n_{.j})}{n} \quad (2)$$

**Dónde:**

$n_i$ : Sumatoria de la fila  $i$  de la matriz observada

$n_j$ : Sumatoria de la columna  $j$  de la matriz observada

$n$ : Sumatoria total de la matriz observada

Tabla N° 27: Matriz de frecuencia esperada (Fe)

5	4	3	2	1	Total
14.25	2.75	2	1	0	20
14.25	2.75	2	1	0	20
14.25	2.75	2	1	0	20
14.25	2.75	2	1	0	20
57	11	8	4	0	80

Fuente: Elaboración Propia (2020)

### 7.3.1.6. Cálculo del chi-cuadrado

A continuación, podemos encontrar los elementos y la Ecuación 3 con la cual podremos obtener el Chi cuadrado calculado (Tabla N° 28).

**Fórmula:**

$$x^2 = \sum_{i=1}^k \frac{(F_{oi} - F_{ei})^2}{F_{ei}} \quad (3)$$

**Dónde:**

Frecuencia Observada = Fo

Frecuencia Esperada = Fe

Tabla N° 28: Matriz de cálculo del chi cuadrado

5	4	3	2	1	Total
0.1096	0.5682	0	0	0	<b>0.68</b>
0.0395	0.2045	0	0	0	<b>0.24</b>
0.0395	0.2045	0	0	0	0.24
0.0044	0.0227	0	0	0	<b>0.03</b>
<b>0.193</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1.19</b>

Fuente: Elaboración Propia (2020)

Tabla N° 29: Matriz de cálculo de P-value

Grados de Libertad	12
X <sup>2</sup> Estadístico Chi-cuadrado	1.19
p-value	1.00
α	0.05
Valor crítico	21.0261

Fuente: Elaboración Propia (2020)

### 7.3.1.7. Prueba de bondad de ajuste

Después de obtener el p-valor (1.00) el cual es mayor  $\alpha$  (0.05) (Tabla N° 29) y el Chi-cuadrado calculado (1.19) el cual se encuentra dentro de la zona de aceptación. Podemos concluir que la hipótesis nula ( $H_0$ ) es aceptada.

### 7.3.2. Prueba de la primera hipótesis

El proceso de extracción, transformación y carga de comentarios desde las diferentes cuentas que las empresas corredoras de seguros tienen en redes sociales permitirá la integración de las diferentes fuentes en un repositorio de datos.

Figura N° 137: Cálculo porcentaje integración extracción de comentarios

```
3 SELECT ROUND(Repositorio/CONVERT(FLOAT,Fuente),2) AS [%Integración]
4 FROM (
5     SELECT COUNT(*) As Fuente
6     FROM STG_FACEBOOK_000
7 ) T1, (
8     SELECT COUNT(*) As Repositorio
9     FROM STG_FACEBOOK_004
10 ) T2
11
12
```



%Integración
0.48

Fuente: Elaboración Propia (2020)

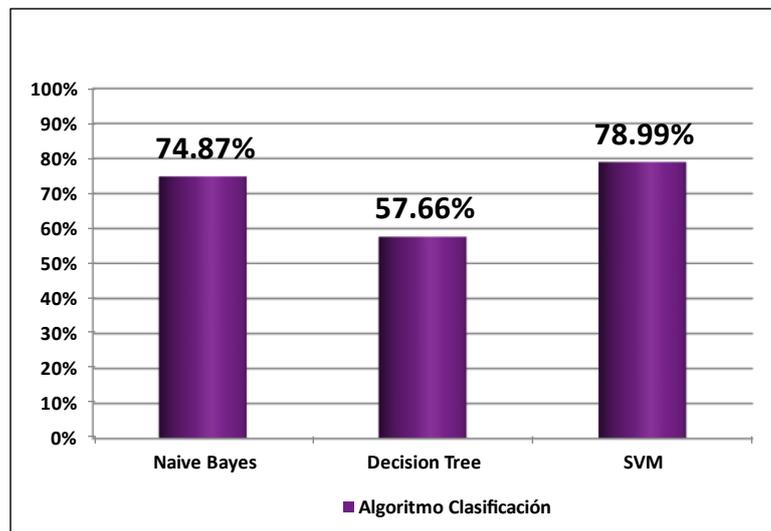
Cómo se puede apreciar en la Figura N° 137, después de ejecutar el proceso de extracción, transformación y carga, el porcentaje de integración de las fuentes es del 48%. el cual es un porcentaje aceptado en nuestro modelo ya que después de realizar la limpieza y descarte de datos

consideramos que si posible lograr la integración de los comentarios en un repositorio. Por lo tanto, consideramos que se cumple a la primera hipótesis.

### 7.3.3. Prueba de la segunda hipótesis

El algoritmo de clasificación de texto que tenga un mejor valor de precisión permitirá la clasificación de un comentario en una categoría.

Figura N° 138: Comparativo precisión algoritmos de clasificación



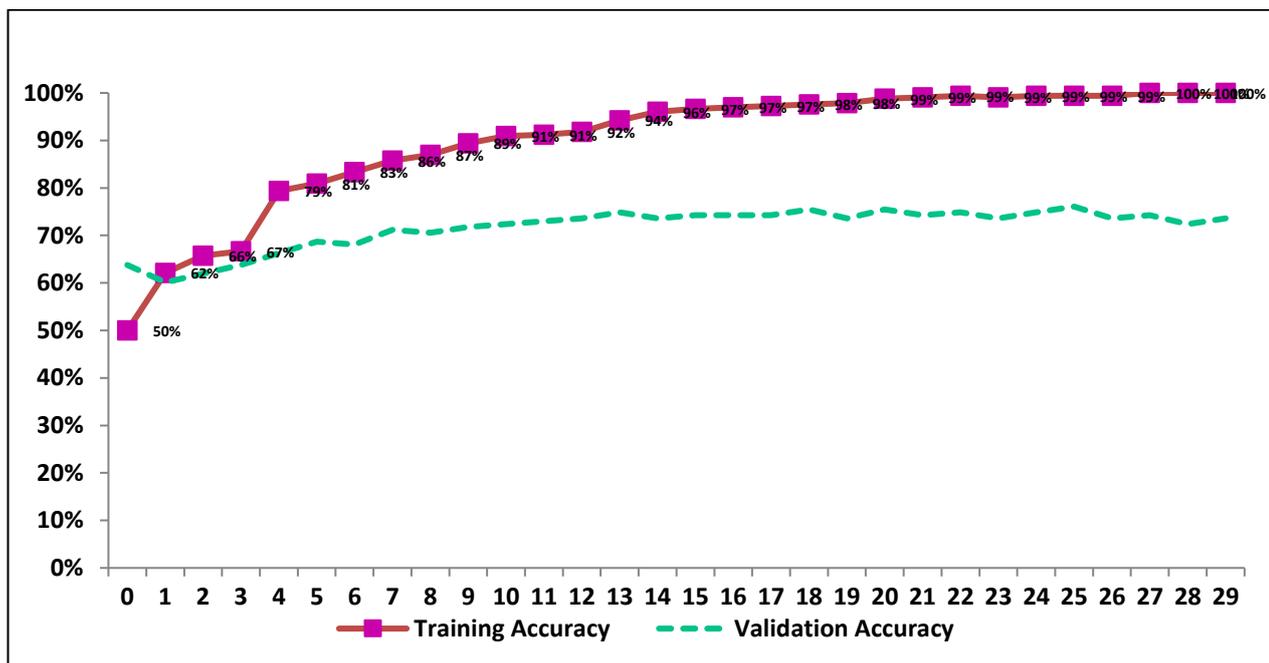
Fuente: Elaboración Propia (2020)

Cómo se puede apreciar en la Figura N° 138, después de comparar el desempeño de los algoritmos de clasificación de texto, el valor más alto en precisión es el SVM con (78.99%). En el desarrollo del modelo de clasificación de texto se ha utilizado el algoritmo de clasificación de texto SVM. Por lo tanto, consideramos que se cumple a la segunda hipótesis.

### 7.3.4. Prueba de la tercera hipótesis

El uso de técnicas de análisis de sentimiento con un alto valor de precisión permitirá identificar las emociones que expresan los usuarios en redes sociales.

Figura N° 139: Porcentaje precisión modelo Análisis de Sentimiento



Fuente: Elaboración Propia (2020)

Cómo se puede apreciar en la Figura N° 139, la precisión del modelo de Análisis de sentimiento tiene una precisión del 76%, lo cual significa que el modelo está en capacidad de predecir en forma acertada si un comentario expresa un sentimiento: positivo, negativo o neutro. Por lo tanto, consideramos que se cumple la tercera hipótesis.

### **7.3.5. Prueba de la cuarta hipótesis**

El diseño de un tablero de control que muestre los indicadores de opiniones y emociones expresadas en redes sociales contribuye a la mejora del servicio al cliente en una empresa corredora de seguros.

#### **7.3.5.1. Hipótesis nula ( $H_0$ )**

El diseño de un tablero de control que muestre los indicadores de opiniones y emociones expresadas en redes sociales contribuye a la mejora del servicio al cliente en una empresa corredora de seguros.

#### **7.3.5.2. Hipótesis alternativa ( $H_a$ )**

El diseño de un tablero de control que muestre los indicadores de opiniones y emociones expresadas en redes sociales no contribuye a la mejora del servicio al cliente en una empresa corredora de seguros.

#### **7.3.5.3. Encuesta realizada**

Para poder probar la cuarta hipótesis se necesita desarrollar la segunda encuesta la cual debe ser llenada por los usuarios del Modelo de Inteligencia de Negocios basado en minería de opinión, la cual podemos encontrar en la Tabla N° 30:

Tabla N° 30: Encuesta 02 – cuarta hipótesis

Pregunta	5	4	3	2	1	Total
La información mostrada en el tablero de control, le ha permitido tomar decisiones acerca de la satisfacción del servicio que tienen los usuarios de seguros?	16	2	1	1	0	20
La información mostrada en el tablero de control, le ha permitido monitorear en forma oportuna cual es el sentir de los usuarios de seguros?	12	3	3	2	0	20
La información mostrada en el tablero de control, le ha permitido responder en forma oportuna a los requerimientos de los usuarios de seguros?	15	3	1	1	0	20
La información obtenida en el tablero de control le ha permitido mejorar del servicio al cliente?	15	4	1	0	0	20

Fuente: Elaboración Propia (2020)

#### 7.3.5.4. Matriz de frecuencia observada ( $F_o$ )

A continuación, podemos encontrar la matriz de frecuencia observada ( $F_o$ ) la cual se puede visualizar en la Tabla N° 31.

Tabla N° 31: Matriz de frecuencia observada ( $F_o$ )

5	4	3	2	1	Total
16	2	1	1	0	20
12	3	3	2	0	20
15	3	1	1	0	20
15	4	1	0	0	20
<b>58</b>	<b>12</b>	<b>6</b>	<b>4</b>	<b>0</b>	<b>80</b>

Fuente: Elaboración Propia (2020)

#### 7.3.5.5. Matriz de frecuencia esperada ( $F_E$ )

Para poder obtener la matriz de frecuencia esperada utilizaremos la Ecuación 4.

**Fórmula:**

$$E_i = \frac{(n_{i.} * n_{.j})}{n} \quad (4)$$

**Dónde:**

$n_i$ : Sumatoria de la fila i de la matriz observada

$n_j$ : Sumatoria de la columna j de la matriz observada

$n$ : Sumatoria total de la matriz observada

Tabla N° 32: Matriz de frecuencia esperada (Fe)

5	4	3	2	1	Total
14.5	3	1.5	1	0	20
14.5	3	1.5	1	0	20
14.5	3	1.5	1	0	20
14.5	3	1.5	1	0	20
58	12	6	4	0	80

Fuente: Elaboración Propia (2020)

### 7.3.5.6. Cálculo del estadístico chi-cuadrado

A continuación, podemos encontrar los elementos y la Ecuación 5 con la cual podemos obtener el estadístico Chi Cuadrado calculado (Tabla N° 33).

**Fórmula:**

$$\chi^2 = \sum_{i=1}^k \frac{(F_{oi} - F_{ei})^2}{F_{ei}} \quad (5)$$

**Dónde:**

Frecuencia Observada = Fo

Frecuencia Esperada = Fe

Tabla N° 33: Matriz de cálculo del chi cuadrado

5	4	3	2	1	Total
0.1552	0.3333	0.1667	0	0	<b>0.66</b>
0.431	0	1.5	1	0	<b>2.93</b>
0.0172	0	0.1667	0	0	0.18
0.0172	0.3333	0.1667	1	0	<b>1.52</b>
<b>0.6207</b>	<b>0.6667</b>	<b>2</b>	<b>2</b>	<b>0</b>	<b>5.29</b>

Fuente: Elaboración Propia (2020)

Tabla N° 34: Matriz de cálculo de P-value

Grados de Libertad	12
X <sup>2</sup> Estadístico Chi-cuadrado	5.29
p-value	0.95
$\alpha$	0.05
Valor crítico	21.0261

Fuente: Elaboración Propia (2020)

### 7.3.5.7. Prueba de bondad de ajuste

Después de obtener el p-valor (0.95) el cual es mayor  $\alpha$  (0.05) (Tabla N° 34) y el Chi-cuadrado calculado (5.29) el cual se encuentra dentro de la zona de aceptación. Podemos concluir que la hipótesis nula ( $H_0$ ) es aceptada.

#### **7.4. DISCUSIÓN DE RESULTADOS**

Los resultados encontrados en la presente investigación fueron obtenidos a través de encuestas realizadas a usuarios del modelo de Inteligencia de Negocios basado en Minería de Opinión, así como también al cálculo del porcentaje de precisión del modelo de Clasificación de Texto para clasificar un comentario en una categoría determinada y la precisión del modelo de Análisis de Sentimiento para predecir si un comentario tiene en un sentimiento: positivo, negativo o neutro.

La precisión del modelo de Clasificación de Texto para asignar a un comentario a una categoría o clase es de 78.99% y la precisión del modelo de Análisis de Sentimiento para transformar un comentario en Sentimiento es de 76.56%, lo cual nos permite concluir que la precisión de los modelos está dentro de un rango adecuado de precisión.

Sin embargo, el modelo de Análisis de Sentimiento no brinda el 100% de precisión para transformar comentarios en sentimiento, pero es una herramienta que puede ser útil para identificar el nivel de satisfacción de los clientes de productos y servicios, sobre todo para poder detectar campañas de desprestigio en redes sociales a tiempo, los cuales deben tratarse de forma inmediata para evitar que la imagen de empresa se vea perjudicada.

Como lo sucedido con la empresa BMW en julio del 2022 tuvo un alza de sus menciones en Twitter, pero las interacciones que expresaban los usuarios eran mayormente negativas por la aparente decisión de la empresa

de vender servicios de suscripción para algunas funciones internas de los carros. El tuit que originó la debacle (Figura N° 129) aglomeró 30,000 retuits y 225,000 me gusta, a tal punto que la empresa BMW se vio obligada a aclarar los planes de suscripción a través de otro tuit (Figura N° 130) y un comunicado oficial.

Ambos modelos de Clasificación de Texto y Análisis de Sentimiento fueron probados y validados por los usuarios del Modelo de Inteligencia de Negocios basado en Minería de Opinión de la empresa, quienes nos dieron sus observaciones para mejorar el funcionamiento y la precisión de los modelos.

Así como también, se realizó la encuesta N° 02 (Anexo 02) a los usuarios del modelo de Inteligencia de Negocios acerca de si los indicadores de opiniones y emociones mostrados en el tablero de control contribuyen a la mejora del servicio al cliente, después de realizar la encuesta y calcular el chi-cuadrado (Tabla N° 34) se pudo comprobar que se obtiene un p-valor de 0.95 que es mayor 0.05 con lo cual se demuestra que el modelo contribuye a la mejora del servicio al cliente.

El Tablero de Control de la presenta investigación fue construido utilizando la herramienta Microsoft Power BI debido a que permite a los usuarios con menor capacidad técnica a modelar, analizar, explorar, administrar datos y compartir hallazgos, asimismo la consultora Gartner ha reconocido a

Microsoft como líder de plataformas de Inteligencia de Negocios y Analítica en su Cuadrante Mágico para el año 2022 (Figura N° 131).

Si bien es cierto los modelos de Clasificación de Texto y Análisis de Sentimiento han sido probados sobre comentarios expresados por usuarios de servicios de productos de banca y seguros, no podemos afirmar fehacientemente que se puedan aplicar fácilmente sobre comentarios o reclamos expresadas en páginas de redes sociales de empresas de otras industrias.

Lo que sí se puede generalizar es la aplicación de la metodología de trabajo utilizado en la presente investigación, así como también la posibilidad de adicionar una quinta capa de Servicios API a los modelos de Inteligencia de Negocio tradicionales los cuales según la metodología Kimball están compuestos de cuatro capas: Fuente de Datos, Extracción Transformación y Carga (ETL), Área de Presentación y Aplicaciones de Inteligencia de Negocios (BI),

La adición de una quinta capa de Servicios API al Modelo de Inteligencia de Negocios (Figura N° 44) abre una inmensa oportunidad de crecimiento en el campo de la Inteligencia de Negocios ya que permite integrar el modelo a grandes bases de datos no estructuradas que se encuentran fuera del alcance de las organizaciones, como son las redes sociales tal como se demuestra en la presente investigación y que en los últimos años han tenido un crecimiento considerable dado a que representan el 80% de los datos

que se generan en el mundo en comparación con su contraparte los datos estructurados los cuales representan solo el 20%.

De la misma manera la quinta capa de Servicios API permite al Modelo de Inteligencia de Negocios tener una arquitectura abierta ya que le permite consumir servicios API, los cuales además pueden estar integrados con modelos de Machine Learning como: Clasificación de Texto, Análisis de Sentimiento, Procesamiento de Imágenes (IP), Reconocimiento automático de la Voz (ASR), Reconocimiento Óptico de Caracteres (OCR), etc. Con esta arquitectura abierta se puede obtener información útil, predecir comportamientos y formas de expresión, tendencias, así como lo expuesto en la presente investigación como es la asignación de un comentario en una categoría determinada y transformar un comentario en un sentimiento positivo, negativo o neutro.

Un modelo de Inteligencia de Negocios con una arquitectura abierta y orientado a servicios API hace que el modelo de Inteligencias de Negocios se enriquezca y que las posibilidades de aplicación del modelo propuesto en la presente investigación se puedan aplicar a otras industrias.

## CONCLUSIONES

Después de desarrollar la presente investigación y de acuerdo con los objetivos planteados se ha llegado a las siguientes conclusiones.

- Las herramientas de desarrollo, algoritmos de clasificación, técnicas de análisis de sentimiento, modelo de datos, modelos de aprendizaje de máquina, modelo de inteligencia de negocios y el enfoque de ingeniería de sistemas se pueden implementar en la presente investigación y en entornos de trabajo reales.
- Para poder seleccionar el algoritmo de clasificación que tenga el mejor valor de precisión. Los algoritmos de clasificación: Naive Bayes, Decisión Tree y SVM (Support Vector Machine) han sido sometidos a una evaluación de su desempeño, resultando que el algoritmo de clasificación SVM con un 78.99% de precisión es el que tiene mejor valor de precisión para ejecutar la clasificación de texto en una categoría. Por lo tanto, el modelo nos ha permitido clasificar los comentarios en una categoría en forma satisfactoria.

- Se ha desarrollado un modelo que utiliza técnicas de análisis de sentimiento, el cual al ser sometido a evaluación de su desempeño nos ha devuelto un 76.07% de precisión para transformar comentarios en información objetiva descubriendo una reacción o sentimiento: positivo, negativo o neutro. Por lo tanto, el modelo nos ha permitido identificar las emociones que expresan los usuarios en sus comentarios en redes sociales.
- Después de realizar la evaluación financiera del proyecto de implementación del modelo de inteligencia de negocios propuesto se ha demostrado que el proyecto es rentable debido a que se obtiene un VAN positivo: 65M y una TIR de 37%
- El uso de tablero de control que muestra indicadores de opiniones y emociones que expresan los usuarios en redes sociales permite contribuir a la mejora del servicio al cliente, la encuesta realizada a los usuarios del modelo y el cálculo del chi-cuadrado de la encuesta retorna que tiene un p-valor igual a 0.95 el cual es mayor al valor de  $\alpha = 0.05$ .

## RECOMENDACIONES

Las recomendaciones están dirigidas a personas o instituciones quienes muestren interés en la aplicación de la investigación o para futuras investigaciones que tomen como base los resultados de la presente investigación.

- Ampliar el ámbito de extracción de comentarios a otras plataformas de redes sociales como: Twitter, Instagram y LinkedIn. Así como a las plataformas para compartir video como Youtube.
- Ampliar la aplicación al campo del Análisis de Sentimiento visual como el procesamiento de imágenes que también son usadas por los usuarios para reaccionar de manera gráfica ante una publicación o comentario y el procesamiento de datos no textuales como son los emoticones, que son una serie de caracteres que representa una cara humana y expresan emociones de alegría, tristeza, amor, etc.
- Agregar información demográfica al proceso de extracción como son: edad, ubicación geográfica, sexo, etc, lo cual permitirá enriquecer el

modelo y el análisis de la información por parte de los usuarios de negocio.

- Integrar el modelo de inteligencia de negocios basado en minería de opinión con las bases de datos transaccionales de la empresa (OLTP), como son: colocaciones, siniestros, reclamos, el cual enriquecerá el modelo y permitirá a los usuarios mejorar la toma de decisiones basados en datos.
- Mejorar la precisión del modelo de Análisis de sentimiento incrementando de manera progresiva los datos de entrenamiento del modelo.
- Escalar el servidor de archivos hacia un entorno de Big Data como Apache Hive, el cual permite procesar grandes cantidades de datos que son almacenados bajo el sistema HDFS y que luego se pueden realizar consultas a los datos usando un lenguaje similar a SQL llamado HiveQL o usando Apache Spark SQL.

## GLOSARIO

- **Inteligencia de Negocios Social:** Es el proceso de extraer datos de las redes sociales, analizarlos para tomar mejores decisiones. Es el enfoque de Inteligencia de negocios enriquecido con "inteligencia social".
- **Minería de Texto:** Es el proceso de explorar y analizar grandes cantidades de datos de texto no estructurados ayudados por un software que puede identificar conceptos, patrones, temas, palabras clave y otros atributos en los datos.
- **Análisis de Sentimiento:** Es el un campo de estudio que analiza las opiniones, sentimientos, evaluaciones, valoraciones, actitudes y emociones de las personas expresados a través de texto.
- **Procesamiento de Imágenes:** es una tecnología apoyada por el uso del computador que es aplicada a las imágenes que nos ayuda a procesar analizar y extraer información útil de ellas.

- **Póliza de Seguros:** Es un contrato independiente en la cual se recogen todas las condiciones generales y particulares que van a regir el contrato de seguro. Es un contrato bilateral que obliga a una de las partes: el asegurado al pago de una prima de seguro y a la otra: la aseguradora a asumir el riesgo correspondiente (indemnización).
- **Prima de Seguro:** Es el precio o costo de un seguro, es el costo que nosotros como asegurados o clientes pagaremos por el uso de la póliza, que obliga a la compañía aseguradora a cumplir con la cobertura.
- **Cobertura de un Seguro:** Es el compromiso que asume una empresa aseguradora de pagar una indemnización al asegurado o a sus beneficiarios con el objetivo de reparar las consecuencias de un siniestro. Tiene un monto límite denominado monto asegurado, el cual se fija al momento de la firma del contrato.
- **Aseguradora:** Es la persona jurídica: sociedad anónima, mutua, cooperativa y mutualidad de previsión social. Cuenta con la autorización oficial correspondiente y asume las consecuencias del riesgo que es objeto de cobertura de seguro.
- **Beneficiario:** Es quien percibe la indemnización en los casos previstos por la póliza. Generalmente tiene un vínculo común de

intereses personales, familiares o económicos con el asegurado o con el tomador del seguro.

- **SQL Server Data Tools (SSDT):** Es una herramienta de desarrollo para crear: modelos de datos de SQL Server Analysis Services (SSAS), paquetes de SQL Server Integration Services (SSIS) y generación de reportes SQL Server Reporting Services (SSRS). Con SSDT se puede diseñar e implementar cualquier tipo de contenido de SQL Server con la misma facilidad con la que desarrollaría una aplicación en Visual Studio.
- **SQL Server Analysis Services (SSAS):** Es la tecnología de Microsoft Business Intelligence para desarrollar soluciones de procesamiento analítico en línea (OLAP). Puede usar SSAS para crear cubos usando los datos para lograr un análisis de datos más profundo y rápido.
- **SQL Server Integration Services (SSIS):** Es una plataforma para crear soluciones de transformación e integración de datos. Se usa para desarrollar soluciones como copiando o descargando archivos, cargando data warehouse, limpiando y extrayendo datos de diferentes fuentes.
- **Emoticonos:** Secuencia de caracteres en notación ASCII que representaba una cara humana y expresaba una emoción. Pueden

expresar alegría u otras emociones positivas se clasifican normalmente como sonrisas.

- **Relational Database Management System (RDBMS):** Sistema de administración de bases de datos (DBMS) que incorpora el modelo de datos relacional, que normalmente incluye una interfaz de programación de aplicaciones de lenguaje de consulta estructurado (SQL). Es un DBMS en el que se organiza la base de datos y se accede a ella de acuerdo con las relaciones entre los elementos. En una base de datos relacional, las relaciones entre elementos de datos se expresan mediante tablas.
- **Base de Datos OLTP:** Son bases de datos que son insertadas, actualizadas, consultadas y mantenidas a través de las transacciones de un sistema de soporte a las operaciones de negocio. Cada transacción involucra registros de bases de datos individuales compuestos por múltiples campos o columnas. Algunos ejemplos que podemos encontrar son los relacionados a la banca, el uso de tarjetas de crédito o el escaneo de productos en cajas de supermercados.
- **Bases de Datos OLAP:** Son bases de datos relacionadas optimizadas para responder a consultas complejas a grandes cantidades de datos históricos, agregados de bases de datos OLTP y otras fuentes, para proyectos de minería de datos, analítica e inteligencia de negocios.

## REFERENCIAS BIBLIOGRÁFICAS

- Abdous, M., & He, W. (2011). Using text mining to uncover students' technology-related problems in live video streaming. *British Journal of Educational Technology*, Vol 42 No 1, 40-49. doi:10.1111/j.1467-8535.2009.00980.x. Obtenido de <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-8535.2009.00980.x>
- Aguinis, H., & Burgi-Tian, J. (2021). Measuring performance during crises and beyond: The Performance Promoter Score. *Business Horizons*, 64, 149e160 doi: <https://doi.org/10.1016/j.bushor.2020.09.001>.
- Alalwan, A., Rana, N., Dwived, Y., & Algharabat, R. (2017). Social media in marketing: A review and analysis of the existing literature. *Telematics and Informatics*, 34(7), 1177–1190. doi:10.1016/j.tele.2017.05.008. Obtenido de <https://www.sciencedirect.com/science/article/abs/pii/S0736585317301077>

APECOSE. (Julio de 2020). *www.apecose.com*. Obtenido de El Corredor de Seguros: <https://apecose.com/el-corredor-de-seguros/>

Arjonilla, A. (2009). El Compromiso de los Clientes como Objetivo Estratégico. *Harvard Deusto Business Review*.

Asociación Peruana de Empresas de Seguros. (2018). *www.apeseg.org.pe*. Obtenido de SEGUROS: UNA VIEJA BUENA IDEA: <https://www.apeseg.org.pe/orientacion/seguros-una-vieja-buena-idea/>

Balbi, S., Misuraca, M., & Scepi, G. (2018). Combining different evaluation systems on social media for measuring user satisfaction. *Information Processing & Management*, Vol 54, Issue 4, Pág 674–685. doi:10.1016/j.ipm.2018.04.009. Obtenido de <https://www.sciencedirect.com/science/article/abs/pii/S0306457317305204>

Batrinca, B., & Treleaven, P. (2015). Social media analytics: a survey of techniques, tools and platforms. *Artificial Intelligence & Society*, 30(1): p. 89–116.

Carli, R., & Paniccia, R. M. (2004). *L'Analisi Emozionale del Testo: Uno strumento psicologico per leggere testi e discorsi*. Milano Italy: FrancoAngeli s.r.l.

Chen, W., & Zou, Y. (2019). Group decision making under generalized fuzzy soft sets and limited. *Engineering Applications of Artificial Intelligence*,

Engineering Applications of Artificial Intelligence 87 (2020) 103344  
<https://doi.org/10.1016/j.engappai.2019.103344>.

Congreso de la República. (Julio de 2011). *congreso.gob.pe*. Obtenido de Normas Legales - Ley de Protección de Datos Personales:  
<http://www.leyes.congreso.gob.pe/Documentos/Leyes/29733.pdf>

Diario Gestión. (2017). *Transformación digital: Seguros*. Obtenido de <https://gestion.pe/blog/innovaciondisrupcion/2015/09/transformacion-digital-seguros.html/?ref=gesr>

Drucker, H., Vapnik, V., & Wu, D. (1999). Automatic text categorization and its applications to text retrieval. *IEEE Transactions on Neural Networks*, 10,5, 1048-1054.

Dumais, S., & Chen, H. (2000). Hierarchical classification of Web content. *SIGIR '00: Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 256–263. <https://doi.org/10.1145/345508.345593>.

Dumais, S., Platt, J., Heckerman, D., & Sahami, M. (1998). Inductive learning algorithms and representations for text categorization. *CIKM '98: Proceedings of the seventh international conference on Information and knowledge management*, pp. 148–155  
<https://doi.org/10.1145/288627.288651>.

Facebook.com. (2019). *Facebook Newsroom*. Obtenido de Facebook Newsroom Stats: [https://newsroom.fb.com/company-info/?\\_ga=2.152025407.72307967.1559368020-579572276.1559368020](https://newsroom.fb.com/company-info/?_ga=2.152025407.72307967.1559368020-579572276.1559368020)

Facebook.com. (s.f.). *Strategic Planning and Operation Lead*. Obtenido de Strategic Planning and Operation LEad: <https://www.facebook.com/careers/jobs/864251053966938/>

Fielding, R. (2000). *Architectural Styles and the Design of Network-based Software Architectures. Doctoral dissertation.*

Forbes.com. (Febrero de 2017). *www.forbes.com*. Obtenido de The Future Of Car Insurance: Digital, Predictive And Usage-Based: <https://www.forbes.com/sites/sarwantsingh/2017/02/24/the-future-of-car-insurance-digital-predictive-and-usage-based/#76eebc2c52fb>

Foucault, M. (1977). *Power/Knowledge - Selected Interview and Other Writings*. NY: Pantheon Books.

Gallinucci, E., Golfarelli, M., & Rizzi, S. (2015). Advanced topic modeling for social business intelligence. *Information Systems*, 53, 87–106. doi:10.1016/j.is.2015.04.005.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137–144. doi:10.1016/j.ijinfomgt.2014.10.007.

Obtenido de

<https://www.sciencedirect.com/science/article/pii/S0268401214001066>

Gestion.com. (Febrero de 2017). *CiberLaw*. Obtenido de Redes Sociales y Privacidad: <https://gestion.pe/blog/cyberlaw/2017/02/redes-sociales-y-privacidad.html?ref=gesr>

Gestion.pe. (2013). *Diario Gestión*. Obtenido de Baja penetración en el uso de inteligencia de negocios: <https://gestion.pe/impresa/baja-penetracion-inteligencia-negocios-28327-noticia/>

Gloor, P. A. (2017). *Sociometrics and human relationships: Analyzing social networks to manage brands, predict trends, and improve organizational performance*. London, UK: Emerald Publishing Limited.

Greco, F. (2016). *Integrare la disabilità. Una metodologia interdisciplinare per leggere il cambiamento culturale*. Milano, IT.

Greco, F., & Polli, A. (2019). Emotional Text Mining: Customer profiling in brand management. *International Journal of Information Management*, doi:10.1016/j.ijinfomgt.2019.04.007. Obtenido de <https://www.sciencedirect.com/science/article/pii/S0268401218313598>

Hajli, N., & Laroche, M. (2019). Applications of business intelligence and analytics in social media marketing. *International Journal of Information Management*, Vol 48 Pág 226-227.

doi:10.1016/j.ijinfomgt.2018.12.014. Obtenido de

<https://www.sciencedirect.com/science/article/pii/S0268401218313422>

Harrysson, M., Metayer, E., & Sarrazin, H. (2019). *How 'social intelligence'*

*can guide decisions*. Obtenido de McKinsey Quarterly:

[https://www.mckinsey.com/industries/technology-media-and-](https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/how-social-intelligence-can-guide-decisions)

[telecommunications/our-insights/how-social-intelligence-can-guide-](https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/how-social-intelligence-can-guide-decisions)

[decisions](https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/how-social-intelligence-can-guide-decisions)

He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text

mining: A case study in the pizza industry. *International Journal of*

*Information Management*, 33(3), 464–472.

doi:10.1016/j.ijinfomgt.2013.01.001. Obtenido de

<https://www.sciencedirect.com/science/article/pii/S0268401213000030>

Jeong, B., Ko, N., Son, C., & Yoon, J. (2020). Trademark-based framework

to uncover business diversification opportunities: Application of deep

link prediction and competitive intelligence analysis. *Computers in*

*Industry*, Volume 124, January 2021, 103356 doi:

10.1016/j.compind.2020.103356.

Jo, T. (2019). *Text Mining Concepts, Implementation, and Big Data*

*Challenge*. Switzerland: Springer Nature.

- Joachims, T. (1999). Transductive Inference for Text Classification using Support Vector Machines. *ICML '99: Proceedings of the Sixteenth International Conference on Machine Learning*, pp 200–209.
- Jones, G., & George, J. (2018). *Essentials of Contemporary Management*. McGraw-Hill Education; Edición: 8.
- Jouve, O., Martin, E., & Guerin, M. (2012). Tutorial M - Case Study: Sentiment-Based Text Analytics to Better Predict Customer Satisfaction and Net Promoter® Score Using IBM®SPSS® Modeler. *Practical Text Mining and Statistical Analysis for Non-structured Text Data Applications*, 509-532 doi: <https://doi.org/10.1016/B978-0-12-386979-1.00021-9>.
- Keiningham, T., Aksoy, L., Cooil, B., & Andreassen, T. (2008). Linking Customer Loyalty to Growth. *MIT Sloan Management Review*, 49. 51-57.
- Killen, C. P., Geraldi, J., & Kock, A. (2020). The role of decision makers' use of visualizations in project portfolio. *International Journal of Project Management*, Volume 38, Issue 5, July 2020, Pages 267-277 <https://doi.org/10.1016/j.ijproman.2020.04.002>.
- Kim Schouten, Rommert Dekker, Flavius Frasinca\*, Rommert Dekker, & Mark Riezebos. (2018). Obtenido de Heracles: A framework for developing and evaluating text mining algorithms: <https://sci->

hub.tw/<https://www.sciencedirect.com/science/article/pii/S0957417419301587>

Kim, W. C., & Mauborgne, R. (2005). *Blue ocean strategy: How to create uncontested market space and make the competition irrelevant.*

Boston, Mass: Harvard Business School Press.

Kimball, R., & Ross, M. (2013). *The Datawarehouse Toolkit , The Definitive Guide to Dimensional Modeling, Third Edition.* Canada: John Wiley & Sons.

Klein, U., & Namjoshi, K. S. (2011). Formalization and Automated Verification of RESTful Behavior. *Computer Aided Verification*, 541–556.

Klinkenberg, R., & Joachims, T. (2000). Detecting concept drift with support vector machines. *ICML '00: Proceedings of the Seventeenth International Conference on Machine Learning*, pp. 487–494.

Kurnia, P. F., & Suharjito. (2018). Business Intelligence Model to Analyze Social Media Information. *Procedia Computer Science*, Vol 135, Pág 5–14. doi:10.1016/j.procs.2018.08.144. Obtenido de Business Intelligence Model to Analyze Social Media Information: <https://www.sciencedirect.com/science/article/pii/S1877050918314303>

Laudon, K. C., & Laudon, J. P. (2014). *Management Information Systems Managing the Digital Firm.* England: Pearson.

- Liu, B. (2015). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge: Cambridge University Press.  
doi:10.1017/CBO9781139084789. Obtenido de Sentiment analysis: Mining opinions, sentiments, and emotions. Sentiment.
- Lovett, M. J., Peres, R., & Shachar, R. (2013). On Brands and Word of Mouth. *Journal of Marketing Research*, 50(4), 427-444.  
doi:10.1509/jmr.11.0458 .
- Lu, Q. S., & Miller, R. (2019). How Social Media Communications Combine with Customer Loyalty Management to Boost Green Retail Sales. *Journal of Interactive Marketing*, Vol 46, 87–100.  
doi:10.1016/j.intmar.2018.12.005. Obtenido de How Social Media Communications Combine with Customer Loyalty Management to Boost Green Retail Sales:  
<https://www.sciencedirect.com/science/article/pii/S1094996818300768>
- Manavi, A. D., & Malini, D. H. (2018). Sentiment Analysis of eWOM on Service Performance of Public. *Amity Journal of Marketing*, 3 (1), (13–23).
- Maron, M. E. (1961). Automatic Indexing: An Experimental Inquiry. *Journal of the Association for Computing Machinery*, 8, 404-417.
- Mertsalov, K., & McCreary, M. (2019). *Document Classification with Support Vector Machines*. Obtenido de Rational Enterprise:

<https://www.rationalenterprise.com/wordpress/wp-content/uploads/2021/04/document-classification-with-support-vector-machines.pdf>

Microsoft. (2022). *Casos de éxito de Power BI*. Obtenido de <https://customers.microsoft.com/en-us/story/1404533828014895899-bayer-creates-cfo-app-for-one-stop-shop-by-steering-relevant-data-using-power-bi>

Mohan Debarchan Mohanty, & Mihir Narayan Mohanty. (2022). *Advanced Data Mining Tools and Methods for Social Computing*. Academic Press.

Mohan Debarchan Mohanty, & Mihir Narayan Mohanty. (2022). *Advanced Data Mining Tools and Methods for Social Computing*. United Kingdom: Academic Press.

Muntean, M., Cabău, L. G., & Rînciog, V. (2014). Social Business Intelligence: A New Perspective for Decision Makers. *Procedia - Social and Behavioral Sciences*, Vol 124, Pág 562 – 567. doi: <https://doi.org/10.1016/j.sbspro.2014.02.520>. Obtenido de <https://www.sciencedirect.com/science/article/pii/S1877042814020680>

Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. Cambridge, Massachusetts: The MIT Press.

National Association Insurance Commissioners. (Mayo de 2019).

*www.naic.org*. Obtenido de TELEMATICS/USAGE-BASED

INSURANCE:

[https://content.naic.org/cipr\\_topics/topic\\_telematicusage\\_based\\_insurance.htm](https://content.naic.org/cipr_topics/topic_telematicusage_based_insurance.htm)

Nishanth, K. J., Ravi, V., Ankaiah, N., & Bose, I. (2012). Soft computing

based imputation and hybrid data and text mining: The case of

predicting the severity of phishing alerts. *Expert Systems with*

*Applications*, 39(12), 10583–10589. doi:10.1016/j.eswa.2012.02.138.

Obtenido de

<https://www.sciencedirect.com/science/article/pii/S0957417412004010>

Osterwalder, A., & Pigneur, Y. (2011). *Generación de Modelos de negocio*.

Barcelona: Grupo Planeta.

Pedrycz, W., & Chen, S.-M. (2016). *Sentiment Analysis and Ontology*

*Engineering Studies in Computational Intelligence*. Springer:

Switzerland.

Power Data. (2017). *Power Data*. Obtenido de 9 estrategias definitivas de

Big Data para el sector Seguros:

<https://cdn2.hubspot.net/hub/239039/file-1958725915->

[pdf/docs/PWD\\_-\\_Ebook\\_-\\_Seguros\\_Big\\_Data-9\\_estrategias\\_-](https://cdn2.hubspot.net/hub/239039/file-1958725915-pdf/docs/PWD_-_Ebook_-_Seguros_Big_Data-9_estrategias_-)

[FINAL\\_v1.pdf](https://cdn2.hubspot.net/hub/239039/file-1958725915-pdf/docs/PWD_-_Ebook_-_Seguros_Big_Data-9_estrategias_-FINAL_v1.pdf)

- Prehofer, C., & Gerostathopoulos, I. (2017). *Modeling RESTful Web of Things Services. Managing the Web of Things, 73-104*. Australia: Morgan Kaufmann.
- RedHat. (2019). Obtenido de Qué son las API y para que sirven?: <https://www.redhat.com/es/topics/api/what-are-application-programming-interfaces>
- Reichheld, F. F. (2003). The one number you need to grow. *arvard Business Review*, 81(12), 46 - 55.
- Salloum, S. A., Al-Emran, M., & Shaalan, K. (2017). Mining Social Media Text: Extracting Knowledge from Facebook. *International Journal of Computing and Digital Systems*, IJCDS Journal 6(2):73-81. doi: 10.12785/ijcnds/060203.
- Sarlan, A., Nadam, C., & Basri, S. (2014). Twitter sentiment analysis. *Proceedings of the 6th International Conference on Information Technology and Multimedia.*, doi:10.1109/icimu.2014.7066632.  
Obtenido de [https://www.researchgate.net/publication/301408174\\_Twitter\\_sentiment\\_analysis](https://www.researchgate.net/publication/301408174_Twitter_sentiment_analysis)
- Sasser, S., Kilgour, M., & Hollebeek, L. D. (2014). *Harnessing the Power of Social Media and Web Analytics.*. USA: IGI Global doi: 10.4018/978-1-4666-5194-4.ch002.

Sebastiani, F. (2002). Machine Learning in Automated Text Categorization.

*ACM Computing Surveys*, Vol. 34, No. 1, March 2002, pp. 1–47.

Sir Francis Bacon. (1668). *Scientia Potestas Est – Knowledge is Power*:

*Francis Bacon to Michel Foucault*. Obtenido de

<https://link.springer.com/article/10.1023/A:1011901104984>

SmartInsights. (2020). *SmartInsights*. Obtenido de Digital Marketing Strategy

Planning Template: <https://www.smartinsights.com/guides/digital-marketing-plan-template/>

Smartinsights. (2020). *Smartinsights.com*. Obtenido de Integrated marketing

megatrends: <https://www.smartinsights.com/advice/integrated-marketing-megatrends/>

Statista.com. (2021). *Social Media & User-Generated Content*. Obtenido de

Statista: <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>

Sullivan, D. (2001). *Document Warehousing and Text Mining: Techniques for*

*Improving Business Operations, Marketing, and Sales*. John Wiley.

Taira, H., & Haruno, M. (1999). Feature selection in SVM text categorization.

*In Proceedings of the 16th Conference on Artificial Intelligence and*

*Eleventh Conference on Innovative Applications of Artificial*

*Intelligence*, pp. 480–486.

- Thangaraj, M., & Sivakami, M. (2018). Text Classification Techniques: A Literature Review. *Interdisciplinary Journal Of Information, Knowledge and Mangement*, 13, 117-135 <https://doi.org/10.28945/4066>.
- Tripathy, A., Agrawal, A., & Rath, S. K. (2016). Classification of Sentiment Reviews using N-gram Machine Learning Approach. *Expert Systems with Applications*, 57 117–126 doi: [dx.doi.org/10.1016/j.eswa.2016.03.028](https://doi.org/10.1016/j.eswa.2016.03.028).
- Tundjungsari, V. (2013). Business Intelligence with Social Media and Data Mining to Support Customer Satisfaction in Telecommunication Industry. *International Journal of Computer Science and Electronics Engineering (IJCSEE)*, Volume 1, Issue 1 (2013) ISSN 2320–4028.
- Vercellis, C. (2009). *Business Intelligence: Data Mining and Optimization for Decision Making*. United Kingdom: John Wiley & Sons Ltd.
- wearesocial.com, & hootsuite.com. (Feb de 2019). *Digital 2019 Peru (January 2019)*. Obtenido de <https://datareportal.com/reports/digital-2019-peru>
- Witten, I. H. (2005). Text mining. *Practical handbook of Internet computing*, 14-1.
- Witten, I. H., & Frank, E. (2005). *Data Mining Practical Machine Learning Tools and Techniques Second Edition*. San Francisco, CA.: The Morgan Kaufmann Series in Data Management Systems.

Yang, Y., & Liu, X. (1999). A re-examination of text categorization methods. *In Proceedings of SIGIR-99, 22nd ACM International Conference on Research and Development in Information Retrieval* , pp. 42-49  
<https://doi.org/10.1145/312624.312647>.

Yoshida, M., Gordon, B., Nakazawa, M., Shibuya, S., & Fujiwara, N. (Febrero de 2018). Bridging the gap between social media and behavioral brand loyalty. *Electronic Commerce Research and Applications*, Vol 28, 208–218. doi:10.1016/j.elerap.2018.02.005. Obtenido de <https://www.sciencedirect.com/science/article/pii/S1567422318300231>

Zephoria.com. (Mayo de 2019). *The Top 20 Valuable Facebook Statistics – Updated May 2019*. Obtenido de <https://zephoria.com/top-15-valuable-facebook-statistics/>

Zephoria.com. (Mayo de 2019). *Zephoria Digital Marketing*. Obtenido de The Top 20 Valuable Facebook Statistics – Updated May 2019: <https://zephoria.com/top-15-valuable-facebook-statistics/>

Anderson, D., Sweeney, D., & Williams, T. (2008). *Estadística Para Administración y Economía* (10a. ed.). México: CENGAGE Learning.

# ANEXOS

## Anexo 01

### Encuesta N° 01 - Hipótesis General

Un modelo de Inteligencia de Negocios basado en Minería de Opinión sobre redes sociales contribuye a la mejora del servicio al cliente en una empresa corredora de seguros en el Perú.

**¿La información obtenida del Modelo de Inteligencia de Negocios basado en minería de opinión sobre redes sociales, le ha permitido tomar decisiones acertadas y oportunas?**

- Nunca
- Raramente
- Ocasionalmente
- Frecuentemente
- Muy Frecuentemente

**¿La información obtenida del Modelo de Inteligencia de Negocios basado en minería de opinión sobre redes sociales, le ha permitido identificar el comportamiento y las opiniones que expresan los usuarios en redes sociales?**

- Nunca
- Raramente
- Ocasionalmente
- Frecuentemente
- Muy Frecuentemente

**¿La información obtenida del Modelo de Inteligencia de Negocios basado en minería de opinión sobre redes sociales le ha permitido identificar problemas en forma proactiva?**

- Nunca
- Raramente
- Ocasionalmente
- Frecuentemente
- Muy Frecuentemente

**¿La información obtenida del Modelo de Inteligencia de Negocios le ha permitido mejorar el servicio al cliente?**

- Nunca
- Raramente
- Ocasionalmente
- Frecuentemente
- Muy Frecuentemente

## Anexo 02

### Encuesta N° 02 – Cuarta Hipótesis

El diseño de un tablero de control que muestre los indicadores de opiniones y emociones expresadas en redes sociales contribuye a la mejora del servicio al cliente en una empresa corredora de seguros.

**¿La información mostrada en el tablero de control, le ha permitido tomar decisiones acerca de la satisfacción del servicio que tienen los usuarios de seguros?**

- Nunca
- Raramente
- Ocasionalmente
- Frecuentemente
- Muy Frecuentemente

**¿ La información mostrada en el tablero de control, le ha permitido monitorear en forma oportuna cual es el sentir de los usuarios de seguros?**

- Nunca
- Raramente
- Ocasionalmente
- Frecuentemente
- Muy Frecuentemente

**¿ La información mostrada en el tablero de control, le ha permitido responder en forma oportuna a los requerimientos de los usuarios de seguros?**

- Nunca
- Raramente
- Ocasionalmente
- Frecuentemente
- Muy Frecuentemente

**¿La información obtenida en el tablero de control le ha permitido mejorar del servicio al cliente?**

- Nunca
- Raramente
- Ocasionalmente
- Frecuentemente
- Muy Frecuentemente

## Anexo 2

### Autorización para la publicación en Acceso Abierto en el Repositorio institucional de la UNI

#### Datos del autor

Nombre y Apellidos: Sander Alex Chauca Parados

DNI / Carné de extranjería / Pasaporte No: 09629789

Correo electrónico: Sander.Chauca@hotmail.com

Teléfono: 998308045

#### Datos del documento:

##### Modalidad de sustentación:

- Tesis
- Trabajo de suficiencia profesional
- Tesina

##### Nombre del grado o título:

Para obtener el grado de maestro en ciencias con mención en ingeniería de sistemas

Ejm: para obtener el grado de bachiller en ciencias con mención en física

##### Nombre del documento:

Modelo de inteligencia de negocios basado en minería de opinión sobre redes sociales para la mejora del servicio al cliente en una empresa corredora de seguros en el Perú

##### Asesor(es):

Mg. Javier Tolentino Conchano Caro

##### Facultad:

Ingeniería Industrial y de Sistemas



Declaración:

Con la presentación de este documento, el (la) autor (a) confirman la originalidad de la obra y que el contenido redactado es producto de su trabajo. Asimismo, garantiza ser el legítimo, único y exclusivo titular de todos los derechos de propiedad intelectual. También, acepta que los contenidos entregados se pueden leer, descargar, reproducir, distribuir, imprimir, buscar y enlazar los textos completos, sin limitación alguna para su posterior registro en el Repositorio Institucional -UNI.

Por lo tanto, Autorizo a la Universidad Nacional de Ingeniería a publicar la obra en el Repositorio Institucional de la UNI avalado por la Ley N° 30035 que regula el Repositorio Nacional Digital de Ciencia, Tecnología e Innovación de Acceso Abierto, su Reglamento (DS N° 006-2015-PCM), así como sus modificatorias, sustitutorias y conexas con el propósito de conservar, preservar y dar acceso abierto a estos recursos.

En consecuencia, la Universidad Nacional de Ingeniería tendrá la posibilidad de divulgar y difundir los contenidos, de manera total o parcial, sin limitación alguna en los medios, canales y plataformas que la Universidad, pudiendo crear y/o extraer los metadatos sobre los contenidos, e incluirlos en los índices y buscadores que estimen necesarios para promover la investigación y el trabajo colaborativo.

Autorizo que el documento sea puesto a disposición del público a través de la siguiente licencia:

Creative Commons Reconocimiento-NoComercial-CompartirIgual 4.0 Internacional. Para ver una copia de esta licencia, visita: <https://creativecommons.org/licenses/by-nc-sa/4.0/>

En señal de conformidad, suscribo el presente documento.



Firma: \_\_\_\_\_

Nombre: Sandro Alex Chauca Parados

DNI / Carné de extranjería / Pasaporte N°: 09629789

Domicilio:

10, 08, 2023  
Fecha AV. Rivara NAVEGRO 548 Dpto. 1010

### Anexo 3

## Hoja de Metadatos complementarios

<b>Datos de autor</b>	
Nombres y Apellidos	Sandro Alex Chavez Parados
DNI o Pasaporte	09629789
ORCID	
<b>Datos de asesor</b>	
Nombres y Apellidos	Javier Tolentino Canchano Caro
DNI o pasaporte	08405192
ORCID	
<b>Presidente del jurado 1</b>	
Nombres y Apellidos	
DNI	
<b>Miembro del Jurado 2</b>	
Nombres y Apellidos	Luis Alberto Lujan Campos
DNI	08076105
<b>Miembro del Jurado 3</b>	
Nombres y Apellidos	Carlos Fidel Ponce Sánchez
DNI	10212510.



Datos de investigación	
Línea de investigación	Inteligencia de negocios aplicando Machine Learning
Grupo de investigación	Si es que no pertenece a un grupo de investigación colocar "No"
Agencia financiadora	Con o sin financiamiento Sin financiamiento
Ubicación geográfica de la investigación	Si se trata de una investigación en laboratorio, universidad, empresa o institución empresa
Año o rango de años en que se realizó la investigación	2020 - 2023
URL de disciplinas OCDE	<a href="https://catalogos.concytec.gob.pe/vocabulario/ocde_ford.html">https://catalogos.concytec.gob.pe/vocabulario/ocde_ford.html</a>

Datos de carácter obligatorio, solicitado por CONCYTEC.



REPUBLICA



DEL PERU

Notaría Rosales Sepúlveda  
Av. Juan de Arona 707 - San Isidro  
Teléfono: 200 - 3700  
secretario@notariarosalessepulveda.com.pe

# UNIVERSIDAD SAN IGNACIO DE LOYOLA

A nombre de la Nación

EL RECTOR DE LA UNIVERSIDAD SAN IGNACIO DE LOYOLA

Confiere el Grado Académico de

*Bachiller en Ingeniería Empresarial y de Sistemas*

a Don (ña):

*Sandro Alex Chauca Paredes*

quien, ha cumplido con los requisitos exigidos por las disposiciones legales vigentes

Por lo tanto, se expide el presente Diploma para que se le reconozca como tal.

Dado y firmado en Lima, el *14* de *Noviembre* de *2011*

*Edward Feber*  
RECTOR



*[Signature]*  
SECRETARIO GENERAL

*Fermin Antonio Rosales Sepúlveda*  
NOTARIO DE LIMA

OFICINA DE REGISTROS Y SERVICIOS ACADÉMICOS  
DIPLOMA N° A01351273 INSCRITO EN EL LIBRO "REGISTRO  
GENERAL DE GRADOS Y TÍTULOS". EN EL ASIENTO  
N° 002558 DE FOJAS 000640, DEL TOMO II  
Lima, 14 de Noviembre de 20 11

ASAMBLEA NACIONAL DE RECTORES



A1351273



Resolución de Rectorado N° 011-2011  
de Fecha: 07/11/11

*Certifico: Que las presentes copias fotostáticas  
que constan de -2- fojas son idénticas  
a la del documento que tengo a la vista.*

05 JUN 2023

*Fernán Antonio Rosales Sepúlveda*  
NOTARIO DE LIMA



**NOTARIA**  
**ROSALES SEPÚLVEDA**

A01351273

REPUBLICA



DEL PERU

Notaría Rosales Sepúlveda  
Av. Juan de Arona 707 - San Isidro  
Teléfono: 200 - 3700  
notario@notariarosalessepulveda.com.pe

# UNIVERSIDAD SAN IGNACIO DE LOYOLA

A nombre de la Nación

EL RECTOR DE LA UNIVERSIDAD SAN IGNACIO DE LOYOLA

Confiere el Título Profesional de

*Ingeniero Empresarial y de Sistemas*

a Don (ña):

*Sandro Alex Chauca Paredes*

quien, después de haber cumplido como alumno (a) de la Universidad con los requisitos exigidos por las disposiciones legales vigentes, optó dicho Título el *15* de *Mayo* de *2014*

POR TANTO: Expide el presente Diploma para que se le reconozca como tal.

Dado y firmado en Lima, el *19* de *Setiembre* de *2014*

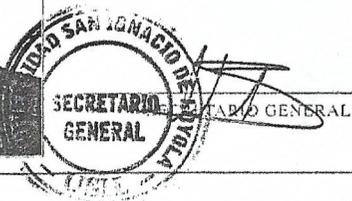
  
RECTOR

*Fernán Antonio Rosales Sepúlveda*  
NOTARIO DE LIMA

  
SECRETARIO GENERAL

OFICINA DE REGISTROS Y SERVICIOS ACADÉMICOS

DIPLOMA Nº A01352106 INSCRITO EN EL LIBRO "REGISTRO GENERAL DE GRADOS Y TÍTULOS". EN EL ASIENTO Nº 737 DE FOJAS 185, DEL TOMO IV  
Lima, 19 de Setiembre de 20 14



ASAMBLEA NACIONAL DE RECTORES



A1352106

Resolución de Rectorado Nº 022-2014 IT  
de Fecha: 08/09/14

*Certifico: Que las presentes copias fotostáticas que constan de -2- fojas son idénticas a la del documento que tengo a la vista.*

05 JUN 2023

*Fernán Antonio Rosales Sepúlveda*  
NOTARIO DE LIMA



**NOTARIA**  
**ROSALES SEPÚLVEDA**

A01352106



**IDIOMAS  
PUCP**

**Notaría Rosales Sepúlveda**  
Av. Juan de Arona 707 - San Isidro  
Teléfono: 200 - 3700  
notaria@notariarosalessepulveda.com.pe

## CERTIFICADO DE NIVEL

### PROGRAMA DE INGLES COMO LENGUA EXTRANJERA CURSO DE INGLES PARA ADULTOS

Idiomas PUCP de la Pontificia Universidad Católica del Perú, otorga el presente certificado a

**Sandro Alex Chauca Paredes**

Quien ha concluido satisfactoriamente sus estudios y cumplido los requisitos de aprobación del **Nivel Intermedio** de este programa en el mes de agosto del año 2015

San Borja, abril 2023



Elsa R. Aragonés Jirón  
Directora Ejecutiva Adjunta



*Certifico: Que esta copia fotostática es idéntica a la del documento que tengo a la vista.*

07 JUN 2023

*Fermin Antonio Rosales Sepúlveda*  
NOTARIO DE LIMA

**NOTARIA**  
**ROSALES SEPÚLVEDA**

Este nivel es equiparable al B2 del Marco Común Europeo de Referencia para las Lenguas (MCER).

**Nota:** En caso este documento sea presentado a un tercero y éste requiera una validación de la información consignada, dirigir su solicitud a [idiomas.chacarilla@pucp.edu.pe](mailto:idiomas.chacarilla@pucp.edu.pe)





# UNIVERSIDAD NACIONAL DE INGENIERÍA

## ESCUELA DE POSGRADO

### CERTIFICADO DE FINALIZACIÓN DE ESTUDIOS DE POSGRADO

FACULTAD: INGENIERÍA INDUSTRIAL Y DE SISTEMAS

MAESTRÍA: EN CIENCIAS

EN: INGENIERÍA DE SISTEMAS

NOMBRES: SANDRO ALEX

APELLIDOS: CHAUCA PAREDES



CÓDIGO: **20176508A**

AÑO DE ADMISIÓN: **2017 - 2**

Nº: **M21140**

CODIGO CURSO	NOMBRE DEL CURSO	CREDITO	NOTA	PERIODO ACADÉMICO
CS-107	ARQUITECTURA DEL COMPUTADOR Y SISTEMAS OPERATIVOS	3	14.2	2017 - 2
CS-207	TEORÍA Y METODOLOGÍA DE SISTEMAS	3	18.0	2017 - 2
CS-307	BASE DE DATOS	3	12.8	2017 - 2
CS-108	REDES Y TELECOMUNICACIONES	3	18.5	2018 - 1
CS-308	INGENIERÍA DEL SOFTWARE	3	15.0	2018 - 1
MA-137	ESTRUCTURAS DISCRETAS EN COMPUTACIÓN	3	13.7	2018 - 1
AS-537	DEONTOLOGÍA Y GESTIÓN AMBIENTAL	2	15.7	2018 - 2
MA-147	INVESTIGACIÓN DE OPERACIONES Y SIMULACIÓN	3	14.2	2018 - 2
TS-207	TÓPICOS ESPECIALES EN INGENIERÍA DE SISTEMAS I (INTELIGENCIA ARTIFICIAL)	3	15.5	2018 - 2
AS-518	PLANIFICACIÓN Y ANÁLISIS DE SISTEMAS	3	15.7	2019 - 1
SE-107	SEMINARIO DE TESIS I	3	14.0	2019 - 1
TS-208	TÓPICOS ESPECIALES EN INGENIERÍA DE SISTEMAS II (COMERCIO ELECTRÓNICO - APLICACIONES INTERNET/INTRANET)	3	17.2	2019 - 1
TS-217	SISTEMAS DE INFORMACIÓN PARA LA TOMA DE DECISIONES (DSS Y SIE/SIG)	3	16.0	2019 - 1
AS-528	CONSULTORÍA Y NEGOCIACIÓN	3	17.5	2019 - 2
GE-468	GESTIÓN DE SISTEMAS DE INFORMACIÓN Y TECNOLOGÍAS DE LA INFORMACIÓN	3	19.5	2019 - 2
PR-707	FORMULACIÓN Y EVALUACIÓN DE PROYECTOS INFORMÁTICOS	3	17.0	2019 - 2
SE-108	SEMINARIO DE TESIS II	3	15.7	2019 - 2

TOTAL DE CREDITOS: **50**

PROMEDIO PONDERADO GENERAL: **15.8**

SISTEMA DE CALIFICACIÓN: De 18.0 a 20.0 = A; De 14.1 a 17.9 = B; De 12.0 a 14.0 = C; Menor a 12 = D

Para optar el Grado de Maestro se requiere haber aprobado el plan de estudios con un promedio ponderado mayor o igual a 14.0, además se deberá sustentar y aprobar la tesis y acreditar dominio de un idioma extranjero o lengua nativa.



Director (e) de la Unidad de Posgrado



Decano



Director (e) de la Escuela de Posgrado

Lima, 06 de Junio de 2023



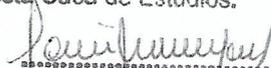
Certifico: Que las presentes copias fotostáticas que constan de 02 fojas son idénticas a la del documento que tengo a la vista.

12 AGO. 2023

  
Fermin Antonio Rosales Sepúlveda  
NOTARIO DE LIMA

**NOTARIA**  
**ROSALES SEPÚLVEDA**



UNIVERSIDAD NACIONAL DE INGENIERÍA  
Secretaría General  
Lima, 15 JUN 2023  
La Secretaría General de la Universidad Nacional de Ingeniería. CERTIFICA : la autenticidad de las firmas precedentes las mismas que pertenecen Dr. Victor A. Cabello Bustamante, Director (e) de la Escuela de Posgrado.  
de esta Casa de Estudios.  
  
M.Sc. SONIA ANAPANULLOA  
Secretaría General





# UNIVERSIDAD NACIONAL DE INGENIERIA

## Oficina de Registro Central y Estadística

### CONSTANCIA DE NO ADEUDOS POSGRADO

Por el presente Certificado se deja constancia que Don (ña):

**CHAUCA PAREDES SANDRO ALEX**

Código U.N.I. : POSG. 20176508A  
Facultad : INGENIERÍA INDUSTRIAL Y DE SISTEMAS  
Especialidad : ING. DE SISTEMAS  
Periodo de egreso : 2019-2

No tiene deudas económicas, ni materiales con la Universidad Nacional de Ingeniería a la fecha. Certifican :

Mag. LUIS ALBERTO ZULOAGA ROTTA  
Decano

Facultad de Ingeniería Industrial Y De Sistemas



Ing. Anita Etelevina PERALDO RAMOS  
Jefe

Oficina Central de Biblioteca



Arq. José Antonio BENLLOCHPIQUER CASTRO  
Jefe

Oficina de Registro Central y Estadística

Por lo que él (la) solicitante se encuentra apto(a), desde el punto de vista económico, para optar el grado académico ó título que solicita.

Lima, 24 de Setiembre del 2021



H- 0013689

El Encargado de la Unidad de Servicios Bibliotecarios y de Información (USBI),  
deja constancia que el egresado: .....Chava Pardo Sando Alex.....  
a la fecha, No.....adeuda libros en la Biblioteca Central.

Lima, 07 de Octubre de 20 21

*p. S. B.*  
Encargado - USBI

001388a

**UNIVERSIDAD NACIONAL DE INGENIERÍA UNI**

Oficina Central de Economía y Finanzas

UNIDAD DE TESORERÍA

DOMICILIO FISCAL:

Av. Túpac Amaru N° 210 - Rimac - Lima - Lima

TELF: 482-5072

R.U.C.: 20169004359

BOLETA DE VENTA  
ELECTRÓNICA

N° B004 - 00256656

SEÑOR(ES) : CHAUCA PAREDES SANDRO ALEX

DNI : 09629789

PRESUPUESTO : ENSEÑANZA POST-GRADO

DEPENDENCIA : FIIS POSTGRADO

FECHA EMISIÓN : 2021-07-25

TIPO MONEDA : SOLES

MEDIO PAGO : Depósito

ÍTEM	CANT.	PARTIDA	DESCRIPCIÓN	UNIDA MEDIDA	PRECIO UNITARIO	VALOR UNITARIO	VALOR DE VENTA
1	1.00	13231421	CONSTANCIA DE NO ADEUDOS POSGRADO	UNI	66.00	66.00	66.00

OP. GRAVADAS	OP. INAFECTAS	OP. EXONERADAS	ANTICIPOS	I.G.V. 18%	TOTAL A PAGAR
0.00	0.00	66.00	0.00	0.00	66.00

SON: SESENTA Y SEIS CON 00/100 SOLES

Incorporado al Régimen de Agentes de Retención de IGV (R.S.135-2002) a partir del 01/11/2002

NOTA:



INFORMACION ADICIONAL
ID ORDEN:90227 ,CONSTANCIA DE NO ADEUDOS - MAESTRIA EN ING. DE SISTEMAS.

Representación impresa de la Boleta de Venta Electrónica  
Podrá ser consultado en <http://www.ocef.uni.edu.pe/webComprobantes>  
Autorizado mediante Resolución de Intendencia N° 0320050000852/SUNAT

ELABORADO POR: admin



**UNIVERSIDAD NACIONAL DE INGENIERÍA UNI**  
**Oficina Central de Economía y Finanzas**  
**UNIDAD DE TESORERÍA**  
DOMICILIO FISCAL:  
Av. Túpac Amaru N° 210 - Rimac - Lima - Lima  
TELF: 482-5072

R.U.C.: 20169004359

**BOLETA DE VENTA  
ELECTRÓNICA**

**N° B004 - 00482710**

**SEÑOR(ES)** : CHAUCA PAREDES SANDRO ALEX  
**DNI** : 09629789  
**PRESUPUESTO** : ENSEÑANZA POST-GRADO  
**DEPENDENCIA** : FIIS POSTGRADO

**FECHA EMISIÓN** : 2023-04-24  
**TIPO MONEDA** : SOLES  
**MEDIO PAGO** : Depósito

ÍTEM	CANT.	PARTIDA	DESCRIPCIÓN	UNIDA MEDIDA	PRECIO UNITARIO	VALOR UNITARIO	VALOR DE VENTA
1	1.00	13231304	GRADO DE MAESTRO	UNI	4,375.00	4,375.00	4,375.00

OP. GRAVADAS	OP. INAFECTAS	OP. EXONERADAS	ANTICIPOS	I.G.V. 18%	TOTAL A PAGAR
0.00	0.00	4,375.00	0.00	0.00	4,375.00

**SON: CUATRO MIL TRESCIENTOS SETENTA Y CINCO CON 00/100 SOLES**

Incorporado al Régimen de Agentes de Retención de IGV (R.S.135-2002) a partir del 01/11/2002

**NOTA:**



Representación impresa de la Boleta de Venta Electrónica  
Podrá ser consultado en <http://www.cef.uni.edu.pe/webComprobantes>  
Autorizado mediante Resolución de Intendencia N° 0320050000852/SUNAT

INFORMACION ADICIONAL
ID ORDEN:401221 ,GRADO DE MAESTRO EN CIENCIAS EN INGENIERIA DE SISTEMAS

ELABORADO POR: admin



UNIVERSIDAD NACIONAL DE INGENIERÍA

Escuela Central de Posgrado

Unidad de Posgrado FIIS

Facultad de Ingeniería Industrial y de Sistemas

**HOJA DE REVISIÓN  
APROBACIÓN DE BORRADOR DE TESIS DE MAESTRÍA**

*Habiéndose presentado y asignado los Asesores de la presente Tesis el 10-02-2021 y concluido con su elaboración; adjunto al presente, remito el Borrador de Tesis, para obtener el Grado de Maestría en Ciencias en Ingeniería de Sistemas, presentado por el Ingeniero:*

**CHAUCA PAREDES SANDRO ALEX**

**Código 20176508A**

Cuyo Título es: **“MODELO DE INTELIGENCIA DE NEGOCIOS BASADO EN MINERÍA DE OPINIÓN SOBRE REDES SOCIALES PARA LA MEJORA DEL SERVICIO AL CLIENTE EN UNA EMPRESA CORREDORA DE SEGUROS EN EL PERÚ”**

*Sírvase informar si es conforme para que se proceda al empaste y la posterior Sustentación de la misma.*

*Lima, 02 de junio de 2023*



**Dr. ERNESTO ROBINSON FLORES CISNEROS**  
*Director (e) de la Unidad de Posgrado FIIS*

**ES CONFORME**

**Asesor: Mg. Javier Tolentino Canchano Caro**

**Jurado Especialista: Dr. Luis Alberto Lujan Campos**

**Jurado Especialista: Dr. Carlos Fidel Ponce Sánchez**

FIIS-UPGM-FOR-013

Av. Túpac Amaru Nº 210 – Rímac, Lima 25, Apartado 1301 – Perú  
Central Telefónica: 481-1070, Anexos: 5210 – 5223

E-mail: [pgfiis@uni.edu.pe](mailto:pgfiis@uni.edu.pe) / [pgfiissecretaria@uni.edu.pe](mailto:pgfiissecretaria@uni.edu.pe) / [pgfiiscoordinacion@uni.edu.pe](mailto:pgfiiscoordinacion@uni.edu.pe)