

UNIVERSIDAD MAYOR DE SAN ANDRÉS
FACULTAD DE CIENCIAS PURAS Y NATURALES
CARRERA DE INFORMÁTICA
POSTGRADO EN INFORMÁTICA



TESIS MAGISTER SCIENTIARUM
PROGRAMA DE MAESTRÍA EN INGENIERÍA DE SOFTWARE
(MAG-SW) 3ra. VERSIÓN GESTIÓN 2007-2008
MODELO DE INTELIGENCIA DE NEGOCIOS PARA
ANALIZAR EL POSICIONAMIENTO DE PRODUCTOS Y
SERVICIOS EN REDES SOCIALES

POR: Ing. Maritza Condori Quispe
TUTOR: M. Sc. José Marcelo Rocabado Torrez

La Paz – Bolivia

2021

DEDICATORIA

Dedico este trabajo de tesis a mis padres quienes siempre han velado por mi bienestar, educación y superación.

AGRADECIMIENTOS

En primer lugar agradezco a Dios por haberme dado toda la fortaleza para continuar y seguir adelante con este proyecto, a mi querida familia por su apoyo constante e incondicional y a mi tutor quien con sus conocimientos y experiencia me guio a través de cada una de las etapas de esta tesis para alcanzar los objetivos establecidos.

RESUMEN

Las redes sociales tales como Facebook y Twitter contienen una gran cantidad de información generada por sus usuarios. Esa información está compuesta de opiniones y comentarios que pueden reflejar tendencias de comportamientos sociales. Se habla de tendencia cuando es posible identificar opiniones y comentarios orientados hacia la misma dirección compartida por una gran cantidad de personas.

Con el presente trabajo se pretende obtener ventajas de los datos generados por los usuarios en las redes sociales Facebook y Twitter, a través de un modelo de Inteligencia de Negocios para analizar el posicionamiento de productos y servicios con los datos provenientes de las redes sociales y para lo cual se abordarán temáticas referidas a las redes sociales, inteligencia de negocios, analítica de negocios, minería de datos, entre otros.

ÍNDICE

INTRODUCCIÓN	1
CAPÍTULO I: ASPECTOS GENERALES	3
1.1 Antecedentes	3
1.2 Planteamiento del problema	4
1.2.1 Formulación del problema de investigación	5
1.2.2 Árbol de problemas	5
1.3 Justificación.....	5
1.4 Objetivo general y específicos de la investigación.....	7
1.4.1 Objetivo general	7
1.4.2 Objetivos específicos.....	7
1.5 Resultados de la Investigación	7
1.6 Operacionalización de variables de la investigación.....	8
CAPÍTULO II: MARCO TEÓRICO	10
2.1 Marco conceptual	10
2.3 Marco Legal	29
CAPÍTULO III: METODOLOGÍA DE LA INVESTIGACIÓN.....	30
3.1 Métodos de Investigación.....	30
3.2 Tipo de investigación	32
3.3 Universo o población de estudio	32
3.3.1 Determinación y elección de la muestra.....	34
3.4 Fuentes y diseño de los instrumentos de relevamiento de información	35
3.4.1 Fuentes de la investigación.....	35
3.4.2 Diseño de los instrumentos de relevamiento de información.....	35
3.5 Procesamiento y análisis de la información	35
CAPÍTULO IV: MARCO PRÁCTICO	36
4.1 Metodología de desarrollo de la solución.....	36
4.1.1 Análisis de Requerimientos.....	37
4.1.2 Modelización (Diseño e implementación del almacén de datos)	54
4.1.3 Desarrollo	63
4.2 Validación de resultados	75

CAPÍTULO V: CONCLUSIONES Y RECOMENDACIONES DE LA INVESTIGACIÓN	81
5.1 Conclusiones y Recomendaciones	81
5.1.1 Conclusiones	81
5.1.2 Recomendaciones	81
5.2 Anexos.....	82
5.3 Bibliografía.....	83

ÍNDICE DE TABLAS

<i>Tabla 1 - Operacionalización de la variable Independiente</i>	<i>9</i>
<i>Tabla 2 - Páginas de Facebook más populares en Bolivia 2021</i>	<i>33</i>
<i>Tabla 3 - Usuarios de Facebook en Bolivia</i>	<i>33</i>
<i>Tabla 4 - Usuarios de Twitter en Bolivia</i>	<i>33</i>
<i>Tabla 5 - Perfiles más populares en Bolivia en Twitter</i>	<i>34</i>
<i>Tabla 6 - Estimación de datos de muestra.....</i>	<i>34</i>
<i>Tabla 7 - Parámetros del Tweet</i>	<i>40</i>
<i>Tabla 8 - Información de los recursos (Tweets)</i>	<i>40</i>
<i>Tabla 9 - Datos comunidad Twitter.....</i>	<i>51</i>
<i>Tabla 10 - Datos del Tweet.....</i>	<i>51</i>
<i>Tabla 11 - Datos del Follower.....</i>	<i>52</i>
<i>Tabla 12 - Datos comunidad Facebook.....</i>	<i>52</i>
<i>Tabla 13 - Datos publicaciones Posts Facebook.....</i>	<i>53</i>
<i>Tabla 14 - Datos Comentarios Facebook.....</i>	<i>53</i>
<i>Tabla 15 - Datos de opiniones Facebook</i>	<i>54</i>
<i>Tabla 16 - Puntuaciones Facebook</i>	<i>54</i>
<i>Tabla 17 - Publicaciones visitantes Facebook</i>	<i>54</i>
<i>Tabla 18 - Detalle de cumplimiento de objetivos</i>	<i>78</i>
<i>Tabla 19 - Cumplimiento de operacionalización de variables.....</i>	<i>80</i>

ÍNDICE DE GRAFICOS

<i>Figura 1 - Árbol de Problemas</i>	5
<i>Figura 2 - Red Social</i>	10
<i>Figura 3 - El panorama de las Redes Sociales</i>	10
<i>Figura 4 - Matriz de adyacencia simétrica</i>	11
<i>Figura 5 - Matriz de adyacencia asimétrica</i>	12
<i>Figura 6 - Publicación en Facebook</i>	14
<i>Figura 7 - Tweet</i>	15
<i>Figura 8 - Esquema tradicional de un proyecto BI</i>	16
<i>Figura 9 - Alcance de una solución BI</i>	20
<i>Figura 10 - Esquema de una solución Social BI</i>	22
<i>Figura 11 - Actividades propias del BA</i>	22
<i>Figura 12 - Dominios de aplicación de BA</i>	23
<i>Figura 13 - Fases de un proyecto de BI</i>	36
<i>Figura 14 - Claves y tokens de acceso a la API de Twitter</i>	42
<i>Figura 15 - Explorador de la API Graph de Facebook</i>	44
<i>Figura 16 - Explorador de la API Graph de Facebook</i>	45
<i>Figura 17 - Modelo conceptual (Twitter)</i>	56
<i>Figura 18 - Modelo conceptual (Facebook)</i>	56
<i>Figura 19 - Procesamiento Extracción, Transformación y Carga</i>	64
<i>Figura 20 - Proceso ETL – Tabla de hechos H_TWEET</i>	67
<i>Figura 21 - Dashboard Facebook I – Vinculación Emocional</i>	69
<i>Figura 22 - Dashboard Facebook II – Vinculación Emocional</i>	69
<i>Figura 23 - Dashboard Facebook Menciones Frecuentes</i>	70
<i>Figura 24 - Dashboard Facebook Menciones Frecuentes</i>	71
<i>Figura 25 - Dashboard Facebook Opiniones</i>	72
<i>Figura 26 - Dashboard Twitter Influencers</i>	73
<i>Figura 27 - Dashboard Twitter Menciones Frecuentes</i>	74
<i>Figura 28 - Dashboard Visibilidad y Exposición en Redes Sociales</i>	74

INTRODUCCIÓN

En la actualidad podemos apreciar que las redes sociales en internet han alcanzado mucha importancia ya que se han convertido en una de las principales herramientas para compartir información, opiniones, gustos e interactuar con personas, empresas, grupos, etc. En los últimos años con la expansión de los dispositivos móviles inteligentes (smartphone) el uso de las redes sociales se ha incrementado considerablemente, es así que podemos decir que han llegado a constituirse en un canal o espacio directo de comunicación que genera bastantes datos e información.

(BARTOLÍ MUÑOZ A., BRAULIO GIL N., CURTO DÍAZ J. & PUIGVERT I PELL F. – 2015, pág. 9). En las redes sociales los usuarios hablan, de forma crítica, de productos, de servicios, de marcas, de empresas, etc. Procesar y analizar esos datos para inferir información sobre la percepción de la empresa, o de sus productos y servicios, por los potenciales clientes llegaría a ser un proceso de alta relevancia para las empresas.

Existen herramientas que permiten analizar los datos de las redes sociales, de hecho las propias redes sociales cuentan con herramientas que brindan información estadística e indicadores acerca del comportamiento de nuestras cuentas en las redes, sin embargo, hay casos donde estos indicadores estadísticos no dan respuesta a las necesidades de información de las empresas, entonces a partir de estos valiosos datos generado en las redes sociales, y con la utilización de técnicas avanzadas para su almacenamiento y análisis, se puede llegar a tener un mejor grado de conocimiento acerca de los clientes o consumidores de la empresa, como por ejemplo:

- Conocer mejor sobre los gustos de los clientes y consumidores sobre productos o servicios de una empresa.
- Segmentar el mercado de manera más eficiente.
- Conocer la apreciación de los clientes hacia la empresa y/o
- Mejorar la atención al cliente.

Dos de las redes sociales con más popularidad en nuestro medio son Facebook¹ que registra más de 7 millones de usuarios activos por mes (FACEBOOK AUDIENCE) y Twitter² donde personalidades e instituciones públicas y privadas registran cuentas activas.

La red social Facebook ofrece variedad de servicios a los usuarios tales como la creación de grupos, mensajería instantánea, video llamadas, creación de páginas, comunidades de amigos, listado de intereses y manejo de eventos. Por su parte la red social Twitter ofrece el servicio de enviar y publicar mensajes de texto cortos (280 caracteres de longitud) en el cual se pueden incluir fotos, links y videos.

Con el presente trabajo se pretende obtener ventajas de los datos generados por los usuarios en las redes sociales Facebook y Twitter, para lo cual se abordarán temáticas referidas a las redes sociales, inteligencia de negocios, analítica de negocios, minería de datos, machine learning, entre otros.

¹ Facebook, red social, <https://www.facebook.com>

² Twitter, red social, <https://twitter.com/>

CAPÍTULO I: ASPECTOS GENERALES

1.1 Antecedentes

En el transcurso de los últimos años cada vez son más las empresas que utilizan soluciones de Inteligencia de Negocios o Business Intelligence BI para la toma de decisiones que mejoren sus resultados empresariales y los productos y servicios ofrecidos a sus clientes. Es así que las empresas se están focalizando en varios aspectos clave que se consideran relevantes en una solución BI.

Por otra parte, durante los últimos años hemos asistido al nacimiento y crecimiento explosivo de las redes sociales. Estas, gracias a su facilidad de uso, han supuesto una revolución en la comunicación y han cambiado de forma significativa los patrones de la comunicación tradicional.

Con el propósito de entender mejor a sus clientes, las empresas requieren incorporar datos de redes sociales en sus sistemas de BI o Inteligencia de Negocio, sin embargo, esta tarea trae cambios importantes a considerar:

- El volumen de información generada por lo general es superior a la que se acostumbra gestionar, en algunos casos puede que sea necesario utilizar nuevos sistemas de almacenamiento.
- El origen y formato de esta información es muy diverso. Con las redes sociales se debe añadir otros tipos de información, como los gustos u opiniones de las personas, que se pueden medir con los “Me gusta”; el impacto de una persona, medible a partir su número de seguidores; la importancia de una información, medible a partir del número de veces que se ha compartido o se ha hecho retuit y el nacimiento de un tema de interés o debate, medible con los trending topics.

Asimismo han surgido muchas herramientas de análisis de medios sociales que facilitan el manejo de esta información a las empresas ANEXO - 2, ya que realizan la obtención y

procesamiento de la información, seguimiento de influencers, monitoreo de cuentas de redes sociales, entre otras.

Sin embargo, la falta de conexión entre la data social con el BI corporativo de la empresa u otros conjuntos de datos puede provocar:

- Una visión incompleta de los factores que afectan al negocio
- Una búsqueda manual y por separado en diferentes sistemas
- Duplicidad de la data.

1.2 Planteamiento del problema

La información que se genera en las redes sociales es enorme e importante para las empresas, hoy en día hay herramientas que nos proveen indicadores estadísticos que nos permiten monitorizar las actividades de las cuentas en las redes sociales, sin embargo, los indicadores no representan información de calidad para las empresas, algunos problemas de información por los que atraviesan las empresas son los siguientes:

- No se tiene una valoración de mis productos y servicios a partir de las opiniones de los consumidores.
- Conocer la reputación de la institución a partir de la información generada en la red social.
- No es posible acceder a una representación o esquema social que sinteticen la información de los consumidores, o de personas interesadas en las actividades de la empresa.
- Se carece de elementos que permitan identificar a líderes de opinión dentro de la red social, es decir, personas que pueden influenciar en la toma de decisiones de aquellas personas que son sus amigos o seguidores.
- No se tiene una visión completa de los factores que afectan al negocio.
- No contar con los datos de las redes genera búsquedas manuales y por separado en diferentes sistemas.

1.2.1 Formulación del problema de investigación

Las herramientas de redes sociales en internet tales como Facebook y Twitter carecen de elementos que permitan extraer información de calidad acerca de los productos y servicios de las empresas la cual les permita guiar mejor sus políticas y estrategias de acción.

1.2.2 Árbol de problemas

Representación del problema a través del siguiente árbol de problemas (ver Figura 1).

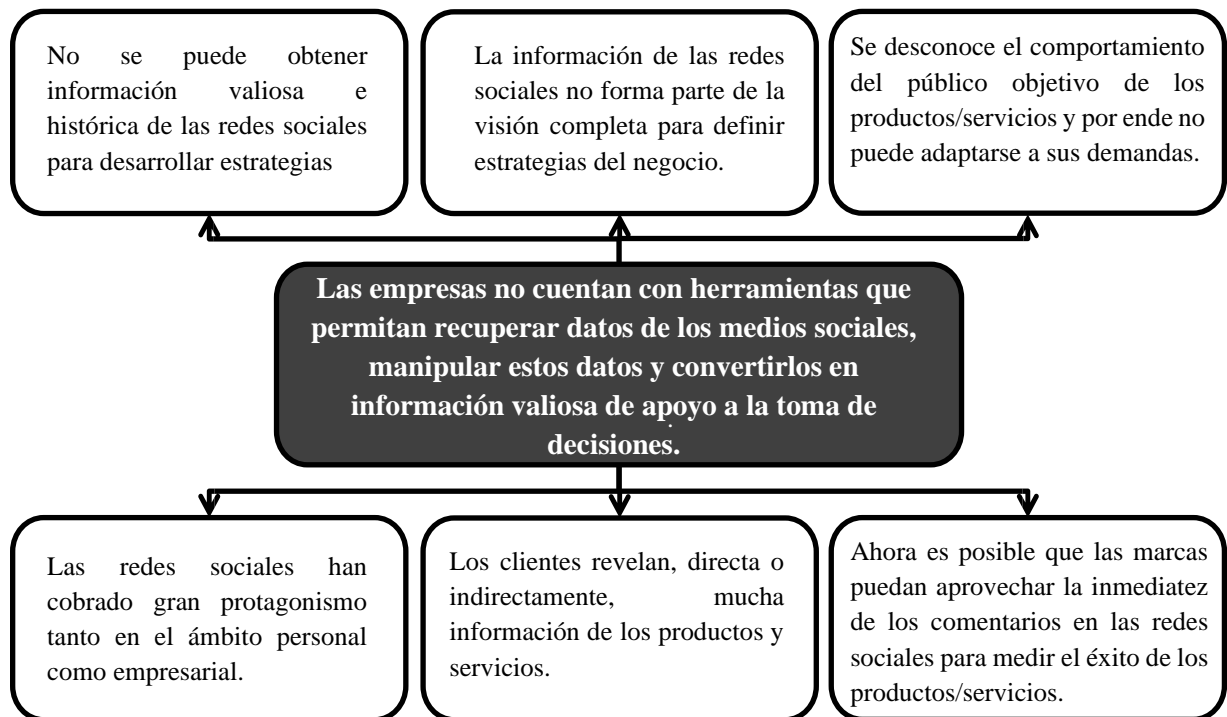


Figura 1 - Árbol de Problemas
Fuente: Propia

1.3 Justificación

A continuación, pasamos a describir la justificación del desarrollo de un modelo de Inteligencia de Negocios con la data proveniente de las redes sociales Facebook y Twitter, se conoce de herramientas que permiten el desarrollo de la analítica social, es decir, que permiten analizar los datos de las redes sociales (Anexo 1), sin embargo, hay casos donde

estos indicadores y herramientas no dan respuesta a las necesidades de las empresas. A saber, de algunas:

- Las herramientas existentes que permiten el análisis de indicadores y métricas son limitados.
- Respecto a los indicadores de rendimiento KPI son específicos no personalizables.
- No permiten el tratamiento y administración de datos históricos.
- Están poco optimizados para altos volúmenes de datos.

Es así que se ha visto que el gran volumen de datos generado en las redes sociales ha traído consigo la necesidad de utilizar nuevos sistemas de almacenamiento como el data warehouse propio de una solución de Inteligencia de Negocios.

Y para el tratamiento de todos los indicadores KPI identificados en cada una de las redes sociales, se ha visto la necesidad realizar el análisis de la información utilizando otras técnicas matemáticas y de minería de textos para elaborar soluciones personalizadas y sacar ventaja de esta información.

Concretamente no se pueden tomar de referencia herramientas de analítica social capaces de satisfacer las siguientes necesidades en su totalidad:

- Capacidad de procesar gran parte de los KPI propios de las redes sociales así como personalizar los mismos.
- Obtener una valoración de mis productos y servicios a partir de las opiniones de los consumidores en las redes sociales.
- Conocer la reputación de la institución a partir de la información generada en la red social.
- Que tengan elementos que permitan identificar a líderes de opinión dentro de la red social, es decir, personas que pueden influenciar en la toma de decisiones de aquellas personas que son sus amigos o seguidores.

- Recuperar automáticamente información de la red social, almacenar en la base de datos y someter la misma al análisis en tiempo real.

1.4 Objetivo general y específicos de la investigación

1.4.1 Objetivo general

Desarrollar un modelo de Inteligencia de Negocios para el soporte a la toma de decisiones en base a información completa y precisa procedente de redes sociales de los consumidores y su valoración hacia los productos y servicios de la empresa.

1.4.2 Objetivos específicos

- Realizar el diseño e implementación del almacén de datos para el almacenamiento de la información proveniente de las redes sociales a través del previo análisis de requerimientos y determinación de tablas de hecho y dimensiones necesarias.
- Desarrollar los procesos de extracción, transformación y carga de datos de las redes sociales al repositorio del modelo de inteligencia de negocios a través de herramientas y técnicas adecuadas.
- Explorar la información generada por el canal de las redes sociales Facebook y Twitter a través de técnicas y algoritmos apropiados como para determinar las preferencias de los consumidores.
- Diseñar tableros de visualización de los datos e información para facilitar la comprensión e interpretación de los datos de las redes sociales.

1.5 Resultados de la Investigación

El modelo de Inteligencia de Negocios que extrae, integra, analiza y explota datos e información de los consumidores proveniente de las redes sociales, genera el conocimiento para el soporte a la toma de decisiones de la empresa.

La toma de decisiones se realiza a partir de información procedente de redes sociales de los consumidores y su valoración hacia los productos y servicios de la empresa.

1.6 Operacionalización de variables de la investigación

En la siguiente tabla (ver Tabla 1) se detalla la operacionalización de variables

VARIABLE	DEFINICIÓN	DIMENSIONES	INDICADORES	ITEMS
<p>VARIABLE INDEPENDIENTE</p> <p>Modelo de Inteligencia de Negocios en redes sociales</p>	<p>La Inteligencia de Negocios, es entendido como el conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización [Conesa & Curto 2010], bajo este mismo concepto un modelo de Inteligencia de Negocios en Redes Sociales constituye la incorporación, procesamiento y exploración de los datos provenientes de las redes sociales a los procesos BI.</p>	<ul style="list-style-type: none"> - Procesos de extracción, transformación y carga de datos (ETL). - Almacén de datos (Datawarehouse) - API de acceso a datos de Twitter y Facebook. - Analítica de datos - Minería de Textos u opinión 	<ul style="list-style-type: none"> - Total seguidores (Twitter) - Nuevos seguidores de perfil - Menciones - Respuestas - Retuits - Total Fans (Facebook) - Total de “Me gusta” en página. - Nuevos “Me gusta” en página (diarios, semanales, mensuales). - Fuente de nuevos “Me gusta”. - Numero de Me gusta en publicaciones. 	<ul style="list-style-type: none"> - Visibilidad y exposición de la empresa - Comunidad (Twitter) - Tweets vinculados a un producto/servicio. - Datos de los seguidores. - Comunidad (Facebook) - Publicaciones (Posts).

VARIABLE	DEFINICIÓN	DIMENSIONES	INDICADORES	ITEMS
<p align="center">VARIABLE DEPENDIENTE</p> <p align="center">Toma de decisiones empresariales</p>	<p align="center">Toma de decisiones empresariales de forma óptima y oportuna.</p>	<ul style="list-style-type: none"> - Análisis de opinión /sentimiento - Herramientas de Visualización de datos. - Cuadros de mando (Dashboards) - Reportes 	<ul style="list-style-type: none"> - Número de “Me gusta” hacia un producto y/o Servicio. - Número de compartidos. - Palabras más usadas para referirse a la empresa en los tweets y/o opiniones. - Palabras positivas. - Palabras negativas. 	<ul style="list-style-type: none"> - Optimización del marketing. - Capturar ideas y clientes insatisfechos. - Cuáles son los más y/o menos favoritos. - Reputación de la institución (Nube de palabras)

Tabla 1 - Operacionalización de la variable Independiente
Fuente: *Elaboración propia*

CAPÍTULO II: MARCO TEÓRICO

2.1 Marco conceptual

2.1.1 Red social

Una red social es una estructura social compuesta por un conjunto de actores y uno o más lazos o relaciones definidos entre ellos (ver Figura 2). Normalmente se representan simbolizando los actores como nodos y las relaciones como líneas que los unen. (WIKIPEDIA Red Social, 2020)



Figura 2 - Red Social

Fuente: https://es.wikipedia.org/wiki/Red_social

En la anterior década el panorama de las redes sociales se había resumido en cuatro categorías: compartir, debatir, establecer contactos y publicar contenidos (Sharing, Discussing, Networking, Publishing). Desde entonces Facebook y Twitter están en el centro del ecosistema de las redes sociales, también se pueden encontrar muchos contendientes en cada categoría (ver Figura 3).



Figura 3 - El panorama de las Redes Sociales

Fuente: <http://www.fredcavazza.net/2013/04/17/social-media-landscape-2013/>

2.1.2 Análisis de redes sociales

- Estructura de un grafo

Un grafo es un conjunto de nodos o vértices y un conjunto de enlaces. Los enlaces pueden considerarse dirigidos o no dirigidos dependiendo de si la relación es mutua o no. Los enlaces entre nodos pueden tener un valor o peso para definir una distancia entre nodos. Los grafos son ampliamente utilizados para encontrar grupos y para la difusión de la información (CUESTA H. & KUMAR S. 2016, pág. 178).

- Grafos no dirigidos

En un grafo no dirigido, no hay distinción entre el origen y el destino de un nodo, lo que significa que la relación entre nodos es mutua, la matriz de adyacencia de un grafo no dirigido es simétrica como se muestra en la siguiente figura (ver Figura 4). Este tipo de grafo es utilizado por la red social **Facebook** donde somos amigos de otros nodos (relación simétrica):

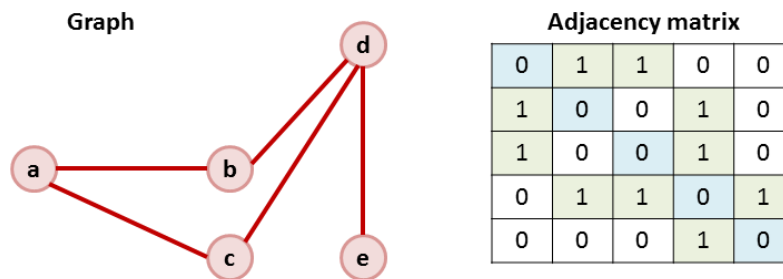


Figura 4 - Matriz de adyacencia simétrica
Fuente: CUESTA H. & KUMAR S. 2016

- Grafos dirigidos

En un grafo dirigido, se establece el origen y el destino de un nodo a través de una fecha, la relación entre nodos es asimétrica o unidireccional, es decir, se tienen dos tipos de relación entre nodos de entrada y de salida (ver Figura 5). Este tipo de grafo es útil para representar la red social **Twitter** donde los nodos entrantes son los Seguidores y los nodos salientes representan a los Seguidos:

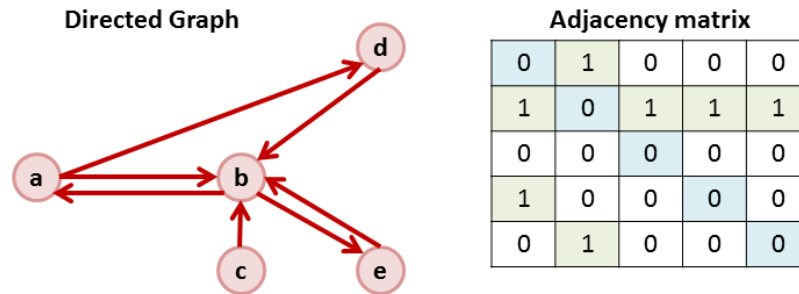


Figura 5 - Matriz de adyacencia asimétrica
Fuente: CUESTA H. & KUMAR S. 2016

2.1.3 Facebook

Facebook es el servicio de red social más conocida y popular de internet, en Bolivia se registra más de 7 millones de personas están activas por mes (FACEBOOK AUDIENCE).

Actualmente Facebook proporciona una variedad de productos y servicios incluidas plataformas de publicidad y comunicaciones a los usuarios (FACEBOOK HELP):

- **Hacer amigos:** Envía solicitudes de amistad a tus amigos, familiares y a otras personas que conozcas y en las que confíes en Facebook.
- **Mensajería:** Los mensajes en Facebook permiten llegar de forma instantánea a las personas que nos interesan, pudiendo enviar fotos, archivos adjuntos, stickers y GIF desde el chat de Facebook o desde Messenger³.
- **Fotos:** Se pueden compartir fotos o crear un álbum de fotos de acontecimientos especiales, pudiendo elegir quien puede ver las mismas. Si alguien te etiqueta en una foto, pero tú no quieres, puedes des etiquetar la foto.
- **Videos:** Se pueden compartir videos pudiendo además editar el mismo y etiquetar a amigos.
- **Páginas:** Las marcas, los negocios, las organizaciones y los personajes públicos pueden usar las páginas para tener presencia en Facebook. Las personas que indican que les gusta una página se convierten seguidores de la empresa.
- **Grupos:** Los grupos ofrecen un espacio para intercambiar opiniones acerca de intereses comunes con determinadas personas. Puedes crear un grupo para cualquier cosa (una

³ **Messenger**, es un servicio de mensajería instantánea y aplicación de software, <https://es-es.messenger.com/>

reunión familiar, el equipo deportivo, etc.) pudiendo personalizar la configuración de privacidad del grupo en función de quién quieres que pueda unirse y ver el grupo.

- **Eventos:** Los eventos te permiten responder a invitaciones y organizar reuniones en el mundo real con personas en Facebook. Se puede crear o asistir a un evento de cualquier clase.
- **Pagos:** Con los pagos de Facebook, puedes comprar en tus tiendas favoritas a través de Facebook, iniciar un grupo de compraventa en tu comunidad y comprar juegos en internet.

(ARIAS LEAL, E. – 2013, pág. 30 y 31) En Facebook los elementos principales a analizar son los recursos y las personas:

- **Recursos.** Son publicaciones con los que los usuarios van a interactuar, pueden ser fotos, preguntas, invitaciones, mensajes, enlaces, vídeos, etc. Todos estos recursos tienen una información en común: comentarios, “me gusta” y número de veces que el recurso ha sido compartido por las personas (ver Figura 6).
- **Personas.** Son los individuos que van a interactuar con los anteriormente comentados recursos. De las personas podríamos tener mucha información, pero dependerá de hasta qué punto la persona nos facilita sus datos. La información personal suele estar abierta salvo que las personas no lo permitan. Al estudiar grandes muestras, tendremos una aproximación de lo que la mayoría tiene en común.

En base a estos elementos de la red social, existen algunos KPI⁴ básicos que pueden utilizarse para monitorizar la actividad de la red social:

- Número de veces que se ha compartido un recurso.
- Recursos que tuvieron más interacciones.
- Número de “Me gusta”.
- Número de comentarios.
- Número de personas que han visto el recurso.
- Localización de las personas.

⁴ **KPI**, indicador clave de rendimiento



Figura 6 - Publicación en Facebook
Fuente: <https://www.facebook.com/HuaweimobileBO>

2.1.4 Twitter

- La anatomía de los datos en Twitter

Twitter es un sitio de Red Social que proporciona un servicio de microblogging⁵ para compartir mensajes de texto de hasta 280 caracteres de longitud denominados tweets, en el cual se pueden incluir fotos, links, videos y otras particularidades de la red social. Podemos recuperar varios tipos de datos de Twitter, tales como Tweets, Seguidores o Followers, Me gusta o Likes, mensajes directos y temas de tendencia o Trends.

- Tweet

Se llama Tweet al mensaje de texto de 280 caracteres. Sin embargo, es posible obtener más información que un simple mensaje de texto (ver Figura 7), también podemos obtener datos tales como: menciones a usuarios (@DZone), hashtags (#BigData, #MachineLearning, #AI y #MachineIntelligence), conteo de retweets, conteo de me gusta, hora y fecha.

⁵ Microblogging: basado en la difusión de mensajes cortos



Figura 7 - Tweet
Fuente: <http://www.twitter.com>

- Seguidores (Followers)

Los usuarios de Twitter pueden seguir a otros usuarios, creando así un grafo dirigido, con varias posibilidades de análisis tales como la centralización y la agrupación de comunidades. En este caso la relación no es mutua por defecto, es así que en Twitter se tiene dos tipos de grados dentro y fuera. Esto puede ser muy útil cuando queremos encontrar al individuo más influyente en un grupo, o que individuo se encuentra entre diferentes grupos.

- Tendencias

Las tendencias en Twitter se expresan con el símbolo “#” antes de una frase o palabra, este símbolo se denomina hashtags y se utiliza específicamente en los Tweets para relacionarlos con un tema, de modo que los usuarios puedan seguir la conversación mediante búsquedas (TWITTER TRENDS). Los temas de tendencia representan un área de análisis de datos en temas como la teoría de difusión de la información.

(ARIAS LEAL, E. – 2013, pág. 31 y 32) Cabe destacar que en Twitter todos los días se vierten millones de opiniones de diferentes asuntos, por lo tanto, analizar sus comentarios puede aportar mucha información para una organización. Los recursos que se pueden analizar en Twitter son:

- **Tweet.** Habiéndose descrito en que consiste, además, cabe destacar que la opinión vertida en un tweet es un elemento que aporta mucha información.
- **Personas.** Cada persona tiene una información asociada básica que puede ser analizada: localización, descripción, nombre, entre otros datos.
- **Hashtag.** Este elemento es muy valioso, mucho más eficiente para el rastreo y búsqueda de opiniones.

2.1.5 Business Intelligence

(CONESA J. & CURTO J. - 2010, pág. 18) Se entiende por Business Intelligence al conjunto de metodologías, aplicaciones prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización. La Figura 8, muestra el esquema tradicional de un proyecto BI y sus principales elementos.

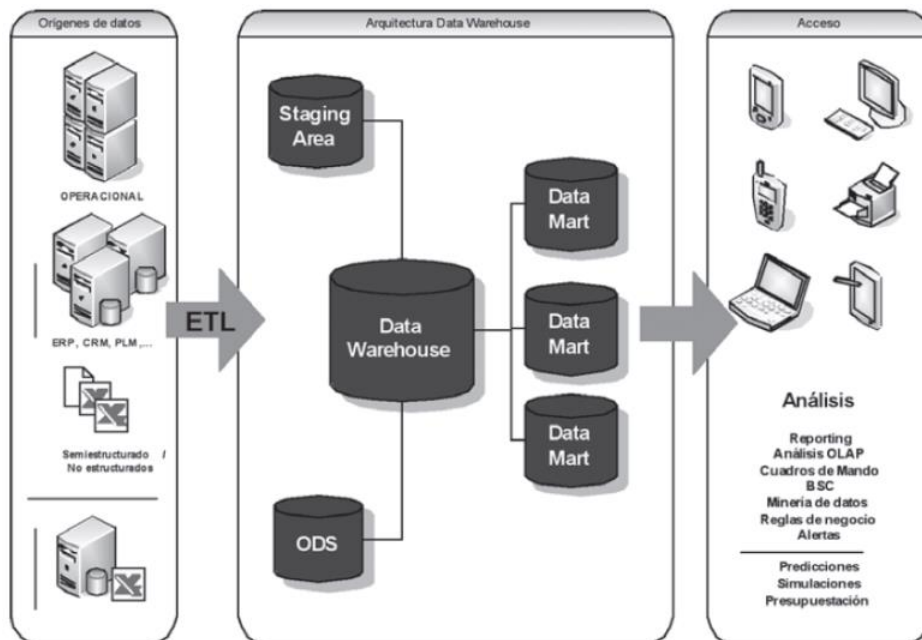


Figura 8 - Esquema tradicional de un proyecto BI
Fuente: CONESA J. & CURTO J. – 2010

- **Orígenes de datos**

Los datos provienen de las bases de datos de los sistemas operacionales o transaccionales de la empresa (estos datos se consideran como la fuente principal de datos internos de la empresa), sistemas de soporte a la información y de fuentes externas a la empresa que proveen información semi-estructurada o no estructurada.

Esta información es sometida a procesos de interacción de datos que permiten extraer, transformar y cargar o también denominados procesos ETL (extract, transform and load) se extrae dicha información y se consolida en una estructura llamada comúnmente data warehouse.

- **Data Warehouse**

Un Data Warehouse es un repositorio de datos que proporciona una visión global, común e integrada de los datos de la organización, dicha información debe ser: estable, homogénea, fiable e histórica, se almacena de forma que permita su análisis desde muy diversas perspectivas, y que a su vez dé unos tiempos de respuesta óptimos. Para ello la información se encuentra altamente desnormalizada y modelada de una forma bastante diferente a los sistemas transaccionales, principalmente se utilizan los modelos en estrella y en copo de nieve.

Un Data Warehouse se caracteriza por ser:

- **Orientado a un tema:** organiza una colección de información alrededor de un tema central.
- **Integrado:** incluye datos de múltiples orígenes y presenta consistencia de datos.
- **Variable en el tiempo o Histórico:** se realizan fotos de los datos basadas en fechas o hechos.
- **No volátil:** sólo de lectura para los usuarios finales.

Un Data Warehouse es comúnmente implementado con la técnica de modelado que conlleva el modelo dimensional, que facilita la comprensión de la base de datos.

- **Modelo Dimensional:** es una técnica de estructura de datos optimizada para herramientas de almacenamiento de datos, su estructura desnormalizada facilita la lectura y el análisis de información, esta técnica es la adecuada para representar datos analíticos. El modelo

dimensional tiene su forma única de almacenamiento de datos y este compuesto por tablas de hechos y dimensiones:

- **Tabla de hechos**, los hechos están compuestos por los detalles del proceso de negocio a analizar, contienen datos numéricos y medidas (métricas) de negocio a analizar, la tabla de hechos (Fact Tables), son tablas que representan dicho proceso de negocio.
 - **Tabla de dimensiones**, Las Dimensiones nos permiten contextualizar los hechos, agregando diferentes perspectivas de análisis a ellos, las tablas de dimensiones son las almacenan la información de las dimensiones.
- **Esquemas del Data Warehouse**
- **Esquema en estrella (star schema)**, cuando el data warehouse se suele modelar utilizando el esquema en estrella, en el hay una única tabla central que es la tabla de hechos, que contiene todas las medidas y una tabla adicional por cada una de las perspectivas desde las que queremos analizar dicha información, es decir por cada una de las dimensiones
 - **Esquema en copo de nieve (snowflake schema)**, se trata de una estructura más compleja que el esquema en estrella. La diferencia es que algunas de las dimensiones no están relacionadas directamente con la tabla de hechos, sino que se relacionan con ella a través de otras dimensiones.

Otros elementos en el contexto de un Data Warehouse son los siguientes:

- **Data Warehousing:** es el proceso de extraer y filtrar datos de los orígenes de datos, para transformarlos, integrarlos y consolidarlos en un Data Warehouse para dar soporte a la toma de decisiones.
- **Data Mart:** es un subconjunto de los datos del Data Warehouse cuyo objetivo es responder a un determinado análisis, función o necesidad, con una población de usuarios específica.
- **Operational Data Store:** es un tipo de repositorio almacén de datos cuyo objetivo es dar respuestas a las necesidades a corto plazo.
- **Staging Area:** es el sistema que permanece entre los orígenes de datos y el Data Warehouse con el objetivo de pre procesar datos complejos y mejorar la calidad de los mismos antes de cargarlos al Data Warehouse.

- **Proceso ETL:** tecnología de integración de datos que permite extraer, transformar y cargar la información al repositorio de datos. También se combina con otras técnicas de consolidación de datos.
- **Metadatos:** datos estructurados y codificados que describen características de instancias; aportan informaciones para ayudar a identificar, descubrir, valorar y administrar las instancias descritas.

Una vez que se tiene la información consolidada es cuando se tiene la posibilidad de hacer análisis de datos de acuerdo a las necesidades empresariales.

- **Herramientas y técnicas de análisis y explotación**

Son las que permiten extraer la información del sistema BI en diferentes estilos según las necesidades o el tipo de información requerida.

- **Reporting:** herramientas para la elaboración de informes, listados y reportes estándar.
- **Monitorización (Cuadro de mando analítico):** elaboración de informes resumen del negocio con gráficos e indicadores clave para la gestión (KPI) muy visuales, que permitan analizar los resultados de la misma de forma rápida y eficaz.
- **Cuadro de mando integral o estratégico (Balanced Scorecard BSC):** elaboración de informes que dan **soporte a la estrategia** de la empresa, ya que ofrecen la visión global de la evolución de los objetivos, iniciativas de actuación e indicadores estratégicos que se organizan en cuatro áreas o perspectivas: financiera, cliente, procesos y formación o desarrollo.
- **Análisis OLAP (on-line analytical processing):** herramientas que manejan interrogaciones complejas en bases de datos relacionales, proporcionando un acceso multidimensional para un análisis profundizando el detalle de los datos.
- **Análisis avanzado/predictivo (minería de datos):** son herramientas que permiten investigar y descubrir patrones ocultos y tendencias presentándolos de forma sencilla y accesible para prever y simular situaciones del negocio.

En función del nivel de complejidad de una solución BI y de las necesidades de la empresa se puede disponer de las herramientas de explotación de la información.

En resumen, una solución de inteligencia de negocio responde a preguntas (ver Figura 9):



*Figura 9 - Alcance de una solución BI
Fuente: Curto J. - 2009*

2.1.6 BI en redes sociales

(BARTOLÍ MUÑOZ A., BRAULIO GIL N., CURTO DÍAZ J. & PUIGVERT I PELL F. – 2015, pág. 7) Durante los últimos años hemos asistido al nacimiento y crecimiento explosivo de las redes sociales. Estas, gracias a su facilidad de uso, han supuesto una revolución en la comunicación y han cambiado de forma significativa los patrones que teníamos de ella. Actualmente las redes sociales forman parte de nuestro día a día: compartimos nuestra vida social y nos entretenemos en Facebook, buscamos trabajo en LinkedIn y nos informamos con Twitter.

Esta nueva forma de comunicación ha traído otros cambios importantes:

- El volumen de información generada suele ser muy superior a la que se gestionaba anteriormente. Eso ha hecho que para poder tratarla eficientemente, en algunos casos sea necesario utilizar nuevos sistemas de almacenamiento, dejando de lado las bases de datos relacionales.
- El origen y formato de esta información es muy diverso. Hasta hace poco la información utilizada en medios *online* era mayoritariamente en formato texto, imagen y vídeo, y estaba relativamente estructurada. Con las redes sociales se debe añadir otro tipo de información, como los gustos u opiniones de las personas, que se pueden medir con los “Me gusta”; el impacto de una persona, medible a partir su número de seguidores; la importancia de una

información, medible a partir del número de veces que se ha compartido o se ha hecho retweet y el nacimiento de un tema de interés o debate, medible con los *trending topics*.

- **Ventajas de las redes sociales para el BI**

(BARTOLÍ MUÑOZ A., BRAULIO GIL N., CURTO DÍAZ J. & PUIGVERT I PELL F. – 2015, pág. 10) La información obtenida de los usuarios de redes sociales es muy heterogénea y permite inferir gran cantidad y diversidad de información.

Al disponer de tanta información relevante, conseguida casi de forma instantánea, los procesos analíticos pueden ser mucho más exhaustivos y la toma de decisiones puede estar respaldada por indicadores más precisos.

La principal ventaja de las redes sociales en los procesos analíticos de una empresa es el alto grado de conocimiento que puede tener de sus clientes, entre otras también se puede destacar:

- Conocer mejor los gustos sobre productos a través de los “Me gusta” en su página de Facebook.
- Obtenemos una estimación sobre los temas de interés a partir de información sobre las personas a las que siguen en Facebook o Twitter.
- Conocemos su grado de satisfacción con una empresa o producto a través de sus comentarios, en blogs, foros, Facebook o Twitter.

Existen también otras ventajas que se derivan del análisis de la información de redes sociales:

- Segmentar el mercado de manera más eficiente
- Mejorar la atención al cliente
- Estudiar a la competencia

Es así que una buena alternativa de análisis de los datos de las redes sociales es a través de las técnicas que proporciona el BI a través del ciclo que recopile los datos de diferentes redes sociales, se los someta a procedimientos de análisis y cree un conjunto de recursos visuales para la explotación de los mismos (ver Figura 10).

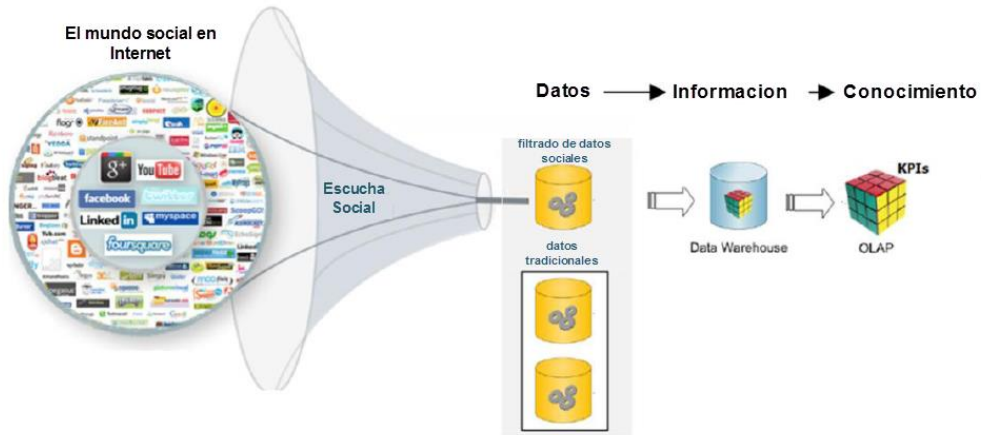


Figura 10 - Esquema de una solución Social BI

Fuente: <http://www.sciencedirect.com/science/article/pii/S1877042814020680>

2.1.7 Business Analytics –BA

(GIRONÉS ROIG J. - 2013, pág. 8-14) Se define por Business Analytics al uso intensivo de datos, de la estadística y del análisis cuantitativo, de los modelos predictivos y explicativos, y de la toma de decisiones basadas en hechos y evidencias. Es la parte de una solución BI centrada en:

La siguiente figura (ver Figura 11) especifica concretamente las actividades propias del BA:

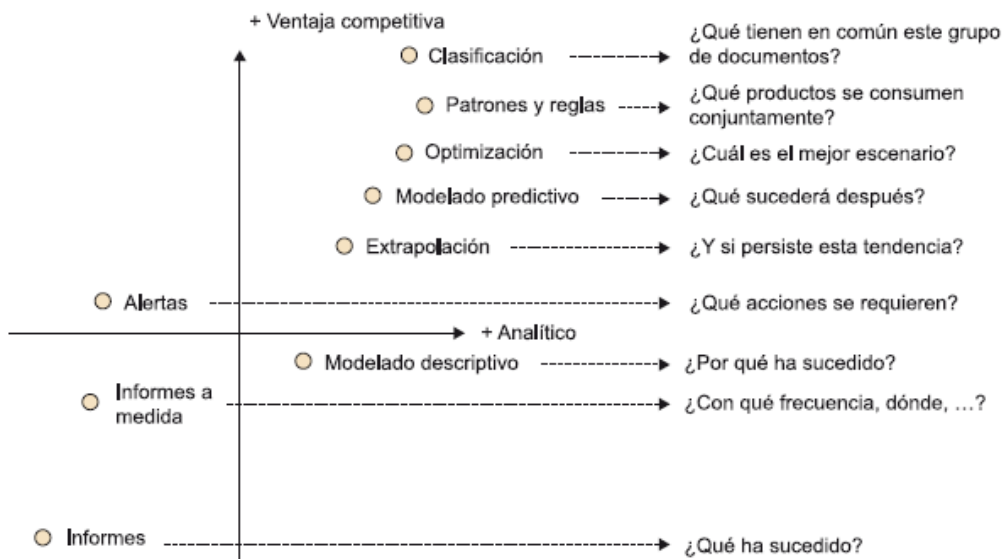
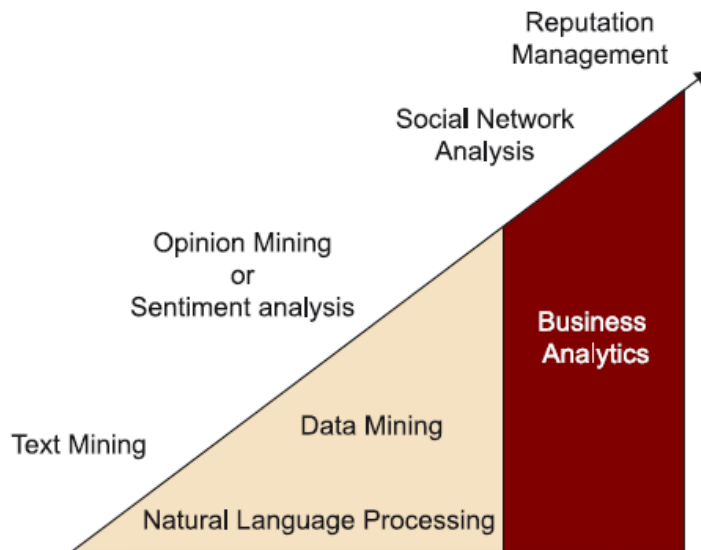


Figura 11 - Actividades propias del BA

Fuente: GIRONÉS ROIG J. – 2013

- Dominios de aplicación de business analytics

(GIRONÉS ROIG J. - 2013, pág. 16) En la siguiente figura (ver Figura 12) se muestran los dominios y paradigmas que aborda la BA para lo cual se sostiene en técnicas propias de la minería de datos o data mining (DM) y de análisis o analytics.



*Figura 12 - Dominios de aplicación de BA
Fuente: GIRONÉS ROIG J. – 2013*

Los distintos paradigmas que han motivado la aparición de las especializaciones de la minería de datos en el campo de la gestión de contenidos, son los siguientes:

- **Text mining (minería de texto):** Actividades específicas sobre un juego de datos no estructurado.
- **Opinion mining o sentiment analysis (análisis de sentimientos):** Estudia la gestión informatizada de texto, que contiene opiniones, sentimientos y subjetividad.
- **Social network analysis (análisis de redes sociales):** Estudia las interacciones y relaciones entre los actores de las redes sociales.

En el marco de la presente investigación, a continuación, describiremos algunas especializaciones:

- **Análisis de sentimientos (Sentiment analysis)**

(WIKIPEDIA_SENTIMENT_ANALYSIS) El análisis de sentimiento (también conocido como minería de opinión) se refiere al uso de procesamiento de lenguaje natural, análisis de texto y lingüística computacional para identificar, extraer cuantificar y estudiar sistemáticamente estados afectivos e información subjetiva.

En términos generales, el análisis de sentimiento intenta determinar la actitud de un interlocutor o un escritor con respecto a algún tema o la polaridad contextual general o reacción emocional a un documento. La actitud puede ser su juicio o evaluación, el estado afectivo (o sea, el estado emocional del autor al momento de escribir), o la intención comunicativa emocional (o sea, el efecto emocional que el autor intenta causar en el lector).

- **Minería de texto (Text mining)**

(WIKIPEDIA TM) Equivalente a la analítica de texto, la minería de texto es el proceso de derivar información de alta calidad a partir de los datos de texto, mediante la aplicación de algoritmos y aprendizaje de patrones estadísticos.

Las tareas típicas de la minería de texto incluyen categorización de texto, agrupamiento de texto, extracción de concepto, producción de taxonomías granulares, análisis de sentimiento, resumen de documento y modelado de relaciones de entidad.

- **Procesamiento de Lenguaje Natural (Natural Language Processing)**

(WIKIPEDIA PLN) Es un campo de las ciencias de la computación, inteligencia artificial y lingüística que estudia las interacciones entre las computadoras y el lenguaje humano. El PLN se ocupa de la formulación e investigación de mecanismos eficaces computacionalmente para la comunicación entre personas y máquinas por medio de lenguajes naturales.

- **Subtareas aplicadas al análisis de sentimientos:**

Una tarea básica en análisis de sentimientos es clasificar la polaridad de un texto dado en un documento, una oración, o en un rasgo o característica, si la opinión expresada en un documento, una oración o un rasgo/característica de una entidad es positiva, negativa, o neutra. La categoría

de sentimientos avanzados, "más allá de la polaridad" parece, por ejemplo, en estados emocionales como "enojado", "triste" y "feliz".

- **Métodos y características:**

Los enfoques existentes para el análisis de sentimiento pueden agruparse en tres categorías principales:

- Las técnicas basadas en el conocimiento, estas técnicas clasifican el texto por categorías de afecto basándose en la presencia de palabras de afecto sin ambigüedad tales como felices, tristes, asustadas y aburridas.
- Los métodos estadísticos, estos sacan ventaja de los elementos del aprendizaje automático (Machine learning), como el análisis semántico latente, las máquinas de vectores de apoyo SVM, la "bolsa de palabras" y la orientación semántica.
- Los enfoques híbridos, aprovechan tanto el aprendizaje mecánico como los elementos de la representación del conocimiento tales como las ontologías y las redes semánticas para detectar semántica que se expresan de una manera sutil.

- **Aprendizaje automático (Machine Learning)**

(WIKIPEDIA MACHINE LEARNING) El aprendizaje automático (Machine Learning) es el subcampo de las ciencias de la computación cuyo objetivo es desarrollar técnicas que permitan a las computadoras aprender sin ser explícitamente programados. Ha evolucionado a partir del estudio del reconocimiento de patrones y la teoría del aprendizaje computacional en la inteligencia artificial.

El aprendizaje automático explora el estudio y la construcción de algoritmos que pueden aprender y hacer predicciones sobre los datos, y tiene una amplia gama de aplicaciones, incluyendo motores de búsqueda, diagnósticos médicos, detección de fraude en el uso de tarjetas de crédito, reconocimiento del habla y del lenguaje escrito, etc.

- **SVM Support Vector Machines**

(WIKIPEDIA SVM) Son un conjunto de algoritmos de aprendizaje supervisado desarrollados por Vladimir Vapnik y su equipo en los laboratorios AT&T.

Estos métodos están propiamente relacionados con problemas de clasificación y regresión. Dado un conjunto de ejemplos de entrenamiento (de muestras) podemos etiquetar las clases y entrenar una SVM para construir un modelo que prediga la clase de una nueva muestra.

Las SVM son capaces de producir buenos modelos para resolver problemas de clasificación binaria, pero también para tareas de regresión, de multi-clasificación y de agrupamiento. Estas propiedades han llevado a las SVM a ser considerados los mejores algoritmos para aplicaciones de clasificación de texto.

2.1.8 Herramientas utilizadas en el desarrollo del modelo

A continuación, describimos las características de las herramientas usadas para el desarrollo del modelo de Inteligencia de Negocios, siendo todas herramientas de acceso libre y/o código abierto.

- **R** es un lenguaje gratuito y libre, basado en módulos e interfaces estadísticos que trabaja con cualquier sistema operativo (Mac, Windows o Linux), su función central es el análisis estadístico y gráfico.

El lenguaje R es un software muy utilizado por la comunidad científica, porque facilita la resolución de problemas estadísticos complejos, proporciona librerías para la recolección de datos, interpretación de la información y graficación.

La principal contribución de R es al campo del análisis estadístico, ciencia de datos, minería y business analytics.

- **Pentaho Data Integration – PDI**, proporciona las capacidades de extracción, transformación y carga (ETL) que facilitan el proceso de captura, limpieza y almacenamiento de datos utilizando un formato uniforme y consistente que es accesible para los usuarios finales.

Pentaho PDI a través de su cliente Spoon permite crear transformaciones y programar y ejecutar trabajos (jobs). Los usos comunes del cliente PDI incluyen: migración de datos entre diferentes

bases de datos y aplicaciones, carga de grandes conjuntos de datos en bases de datos aprovechando al máximo los entornos de procesamiento en la nube, limpieza de datos con pasos que van desde transformaciones muy simples hasta muy complejas entre otras.

- **Power BI**, es un servicio de análisis empresarial de Microsoft, su objetivo es proporcionar visualizaciones interactivas y capacidades de inteligencia empresarial (business intelligence) con una interfaz lo suficientemente simple como para que los usuarios finales creen sus propios informes y paneles.
- **PostgreSQL** es un gestor de base de datos relacionales open source, multiplataforma, es decir que puede ser usado en Linux, Unix o Windows. Marco Referencial

Las ventajas principales que presenta PostgreSQL son: estabilidad, escalabilidad y configuración, estándar SQL, robustez y fiabilidad y soporte.

En cuanto a los límites prácticos de almacenamiento de datos, estos están sujetos a las limitaciones del motor de la base de datos y el espacio disponible en disco donde se ejecuta. Sin embargo, también es importante considerar las capacidades y límites rígidos propios del motor⁶.

2.2 Marco Referencial

La inteligencia de negocios ha cobrado importancia en los últimos años y más aún con las redes sociales son un tema relevante en el contexto académico, se ha encontrado las siguientes investigaciones referenciales al Social Business Intelligence:

(DINTER, B. & LORENZ, A., 2012) Tanto los dominios de la Inteligencia de Negocios (BI) y los medios sociales actualmente se han convertido en importantes campos de investigación. Mientras que una solución BI tiene como objetivo apoyar a la toma de decisiones empresariales al proporcionar datos analíticos relevantes, las redes sociales son una fuente emergente de conocimiento personal e individual, opiniones y actitudes de los interesados. De un tiempo hasta esta parte se observa una convergencia de los dos dominios en las implementaciones e investigaciones del mundo real, resultando un concepto como el Social BI. Varias cuestiones de investigación siguen abiertas en este contexto, en ese marco, la publicación científica propone

⁶ PostgreSQL Limits, <https://www.postgresql.org/docs/12/limits.html>

una agenda de investigación para un Social BI, contrasta las características de las redes sociales con las áreas de diseño de BI, de las cuales las se tomaron en consideración las siguientes:

- **Productos y servicios:** con la disponibilidad de los datos de las redes sociales, se pueden ofrecer productos y servicios de BI nuevos o ampliados. La investigación está en cómo incluir datos de redes sociales y resultados de análisis (como estructuras de red, resultados de análisis de sentimiento, etc.) en productos de BI.
- **Datos:** los datos de las redes sociales, conlleva a realizar las tareas de recolección, procesamiento e integración de datos para BI social de manera diferente, lo cual pueden servir como un punto de partida apropiado para futuras investigaciones en este tema.
- **Tecnologías de información y comunicación:** los altos volúmenes de datos y las tasas de actualización frecuentes de los datos de las redes sociales tienen un impacto en las TIC. Sorprendentemente, hay muy pocas contribuciones científicas disponibles que brinden el apoyo adecuado. Los datos no estructurados requieren un software para el procesamiento de datos. El concepto muy popular actualmente de "big data" debería ofrecer soporte para esta área de diseño de BI.
- **Técnicas:** características propias de las redes sociales (alto volumen de datos y datos no estructurados, también contenido nuevo provisto por las redes sociales) resultan en una demanda de investigación considerable para las técnicas de análisis, por ejemplo, otras técnicas, como ETL, no se abordan hasta ahora.

(ASENSIO B., E. – 2015). Partiendo de que las técnicas de minería de datos permiten obtener información de redes sociales como Twitter, se destaca la aplicación de minería de datos en redes sociales/web, más concretamente técnicas de minería de datos aplicado al análisis lingüístico de los textos de Twitter.

(FRANCIA, M., GOLFARELLI, M. & RIZZI, S. - 2014). Considerando al Social BI como una disciplina emergente que tiene como objetivo combinar datos corporativos con contenido textual generado por el usuario para permitir que los responsables de la toma de decisiones analicen su negocio en función de las tendencias percibidas del entorno, todavía no hay una metodología de diseño específico y orgánico disponible. Se propone una metodología iterativa para el diseño y mantenimiento de aplicaciones Social BI.

(MUNTEAN, M. CABĂU, L. & RÎNCIOG, V. - 2014). Basado en recientes contribuciones de autores, se introduce el concepto de Social Business Intelligence, a través de un marco teórico general para definir y modelar la perspectiva social, como un complemento al esquema del repositorio datos del BI clásico, considerando que los datos de las redes sociales definen una nueva vista para los responsables de toma de decisiones.

2.3 Marco Legal

(ARTICULO XIII) En Bolivia no existe una regulación específica para los medios de información que tienen sus espacios en Internet. Para los medios impresos rige la Ley de Imprenta, vigente desde el 19 de enero de 1925.

Tampoco existe normativa que restrinja las opiniones en Internet, aunque la organización sindical cocaleros de la región del Chapare, propuso regular las opiniones que se difunden a través de las redes sociales. La demanda surgió días después del 21 de febrero, fecha en que el presidente Evo Morales y su vicepresidente, Álvaro García Linera, perdieron en una consulta popular cuando el 51,3 por ciento de los electores rechazaron una modificación constitucional y su cuarta postulación.

En opinión de Morales y las organizaciones sociales afines al Movimiento al Socialismo MAS, las redes sociales como Facebook y Twitter, generaron una corriente de rechazo al pedido de modificación constitucional y determinaron su derrota en las urnas, la primera en 10 años en que gozó de popularidad para ser reelegido en tres oportunidades.

No existe tampoco normativa que exija la veracidad de información sobre opiniones vertidas en redes sociales. Sin embargo, cada medio informativo instalado como empresa periodística, posee un manual de funciones, normas internas y un permanente proceso de análisis de la veracidad de la información a cargo de sus periodistas.

CAPÍTULO III: METODOLOGÍA DE LA INVESTIGACIÓN

3.1 Métodos de Investigación

En esta etapa, se señalan las actividades que serán realizadas en el desarrollo del trabajo de investigación con lo cual se podrá obtener los objetivos de la solución propuesta:

3.1.1 Fases metodológicas

A continuación, se detallan las fases metodológicas que serán aplicadas en el presente trabajo de investigación hasta llegar a la comprobación de la hipótesis:

Paso 1. Planteamiento del problema cuantitativo

En este paso, se busca conocer la problemática que impulsa el desarrollo de la investigación, para esto se tomarán en cuenta los siguientes aspectos:

1. Propósito del proyecto y justificación de negocio.
2. Que incluirá el proyecto de investigación/ Que NO incluye el proyecto.
3. Objetivos del proyecto/ Que se pretende con la implantación de la solución.
4. Definición de los hitos significativos y actividades.

Paso 2. Planificación

En este paso se determinará qué hay que hacer, en qué tiempo y con qué recursos para cumplir un objetivo, por lo que en este paso se determinarán los siguientes aspectos:

1. Un cronograma que establece las actividades que hay que llevar a cabo para alcanzar los objetivos.
2. Una definición de los tiempos y recursos necesarios para completar el trabajo.
3. Un mecanismo para monitorear avances, controlar el alcance y gestionar el proyecto para asegurar los resultados finales.

Paso 3. Desarrollo de la Perspectiva Teórica

En esta etapa se contextualiza el problema de investigación, se hará la revisión de la teoría concerniente a:

1. Solución Business Intelligence: Ciclo de vida de la información, fases estándar metodologías, componentes y herramientas.
2. Analítica de datos e información: Métricas e indicadores, algoritmos y técnicas para el tratamiento de datos no estructurados o semiestructurados.

Paso 4. Diseño e implementación de un modelo de Business Intelligence para analizar el posicionamiento de productos y servicios.

Una vez que se cuente con todos los insumos para el desarrollo de la investigación, se procederá con el desarrollo del proyecto objetivo. Es así que la ejecución puede sostenerse sobre los siguientes aspectos:

1. Identificar la arquitectura adecuada para una solución Social Business Intelligence.
2. Identificar los pasos para la construcción de un social Business Intelligence.
3. Identificar las metodologías y herramientas que serán para el desarrollo de la solución.

Paso 4. Probar la Hipótesis.

Con el prototipo desarrollado en la fase anterior, se procederá a realizar el análisis estadístico de acuerdo a los KPI específicos y elementos propios del Social Business Intelligence.

3.1.2 Técnicas de investigación

Como técnica de investigación y para la recolección de datos apropiados para el análisis, se hará uso de encuesta apoyada por un cuestionario realizada al personal (técnicos, directivos y/o alta dirección) que toma decisiones de diferente alcance, pero en relación a la información y datos provenientes del canal de las redes sociales, con el objetivo de conocer el nivel de madurez de aplicabilidad y uso de los servicios de redes sociales para la toma de decisiones del negocio. Con lo cual se busca conocer las necesidades más urgentes y necesarias de información a ser obtenidas de estos medios.

3.2 Tipo de investigación

De acuerdo con las características de la investigación a desarrollar, se considera el siguiente tipo de investigación:

Cuantitativa - Cualitativa

Es **Cuantitativa - Cualitativa**, porque la comprobación de la hipótesis es en base a datos e información que serán interpretados a partir de los resultados obtenidos del modelo de Inteligencia de Negocios.

3.3 Universo o población de estudio

El universo de estudio de la presente investigación corresponde a:

- Empresas y negocios que han optado por contar con una cuenta en las redes sociales Facebook y Twitter a efectos de proporcionar información sobre productos y servicios, realización de campañas de marketing, mejorar el servicio de atención al cliente y otros.

Según datos estadísticos oficiales de <https://www.socialbakers.com>, empresa que realiza los análisis y marketing en redes sociales para muchas marcas en el mundo; para Bolivia se tienen los siguientes datos:

Red social Facebook:

Tabla 2: Páginas de empresas de productos y servicios más populares en Bolivia en la categoría marcas de empresas y negocios.

Nro	Empresa	Número de fans
1.	Tigo Bolivia	1.970.191
2.	Viva	1.401.888
3.	Casa Elena	975.972
4.	Paceña	869.900
5.	Dismac	699.477

Nro	Empresa	Número de fans
6.	Entel Bolivia	686.269
7.	Banco Nacional de Bolivia S.A.	685.545
8.	Mi Teleférico	571.685
9.	Banco Económico	550.224
10.	Fair Play	529.879

*Tabla 2 - Páginas de Facebook más populares en Bolivia 2021
Fuente: [https://www.socialbakers.com]*

Tabla 3: Facebook presenta un total de 7.7 millones de usuarios activos en Bolivia. El crecimiento medio de esta plataforma en Bolivia es del 3% trimestral.

Usuarios	Mujeres	Hombres
7.7M	47.4%	52.6%

*Tabla 3 - Usuarios de Facebook en Bolivia
Fuente: [https://cecalizaya.com/redes-sociales-en-bolivia]*

Tabla 4: Twitter presenta un total 318.000 usuarios activos en Bolivia.

Usuarios	Mujeres	Hombres
318.000	34.6 %	65.4 %

*Tabla 4 - Usuarios de Twitter en Bolivia
Fuente: [https://cecalizaya.com/redes-sociales-en-bolivia]*

Tabla 5: Perfiles de negocios más populares en Bolivia según <https://www.socialbakers.com/>

Nro	Empresa	Número de seguidores
1.	BoABolivia	46.825
2.	Huaweimobilebo	32.665
3.	ENTEL S.A.	30.711
4.	Tigo Bolivia	23.594
5.	YPFB Corp.	23.046

Nro	Empresa	Número de seguidores
6.	VIVA	16.675
7.	Coca-Cola Bolivia	15.411
8.	Cedib	14.707
9.	ATT Bolivia	14.015
10.	Cerveza Paceaña	11.123

*Tabla 5 - Perfiles más populares en Bolivia en Twitter
Fuente: [https://www.socialbakers.com/]*

3.3.1 Determinación y elección de la muestra

Para el caso práctico se eligió a la empresa de servicios Tigo⁷, la cual tiene una presencia importante en las redes sociales Facebook y Twitter (ver Tabla 6) a efectos de proporcionar información sobre productos y servicios, realización de campañas de marketing, mejorar el servicio de atención al cliente, aumento de la visibilidad e imagen de la marca, facilitar la comunicación con los clientes, promoción de servicios y otros.

Actualmente tiene la Página de Facebook (Fanpage) con mayor audiencia de Bolivia entre las páginas de empresas y una presencia importante en Twitter.

Datos en redes sociales:

Empresa	Número de Fans en Facebook	Número de Seguidores en Twitter
TIGO BOLIVIA	1.970.191	23.575

*Tabla 6 - Estimación de datos de muestra
Fuente: Elaboración propia*

Es así que se considera disponer de una gran cantidad de información relevante, heterogénea y variada obtenida de forma instantánea y a un costo muy reducido, la cual nos permita realizar todos los procesos analíticos planteados para el cumplimiento de los objetivos.

⁷ Tigo Bolivia, es una empresa de telecomunicaciones boliviana, ofrece servicios de telefonía móvil, internet móvil y televisión satelital, [https://es.wikipedia.org/wiki/Tigo_\(Bolivia\)](https://es.wikipedia.org/wiki/Tigo_(Bolivia)).

3.4 Fuentes y diseño de los instrumentos de relevamiento de información

Las redes sociales permiten recuperar parte de su información con el uso de plugins, widgets o consultas a sus APIs disponibles.

Las API para las redes sociales son bibliotecas informáticas que permiten a un programa externo, través de consultas a la API, extraer datos de la red social para su posterior análisis y generación de información. Estas consultas se pueden personalizar mediante parámetros, como el idioma de los mensajes, la fecha, región de publicación, entre otros.

3.4.1 Fuentes de la investigación

Las fuentes primarias de investigación son las propias redes sociales ya que contienen la información central que será el objeto de estudio y análisis en el presente trabajo.

Las fuentes secundarias serán aquellas referentes a las herramientas que emplearemos y que nos permitirán realizar el desarrollo del trabajo de investigación.

3.4.2 Diseño de los instrumentos de relevamiento de información

No se ha procedido al diseño de instrumentos propios de relevamiento de información de las redes para la realización de pruebas al modelo desarrollado, debido a que las redes sociales proporcionan bibliotecas y/o herramientas con las cuales es posible la obtención de la información y datos.

La recuperación de datos puede realizarse mediante librerías de R o Python, en nuestro caso usaremos librerías de R para realizar una búsqueda de datos en las redes sociales.

3.5 Procesamiento y análisis de la información

El procesamiento y análisis de los datos e información se encuentra detallado en el siguiente capítulo Marco Practico.

CAPÍTULO IV: MARCO PRÁCTICO

4.1 Metodología de desarrollo de la solución

Para el desarrollo del Modelo de Inteligencia de Negocios para analizar el posicionamiento de productos y servicios en Redes Sociales, seguiremos las tres primeras fases definidas para el desarrollo de un proyecto y/o solución BI (ver Figura 13).

- Fases del desarrollo de una solución BI

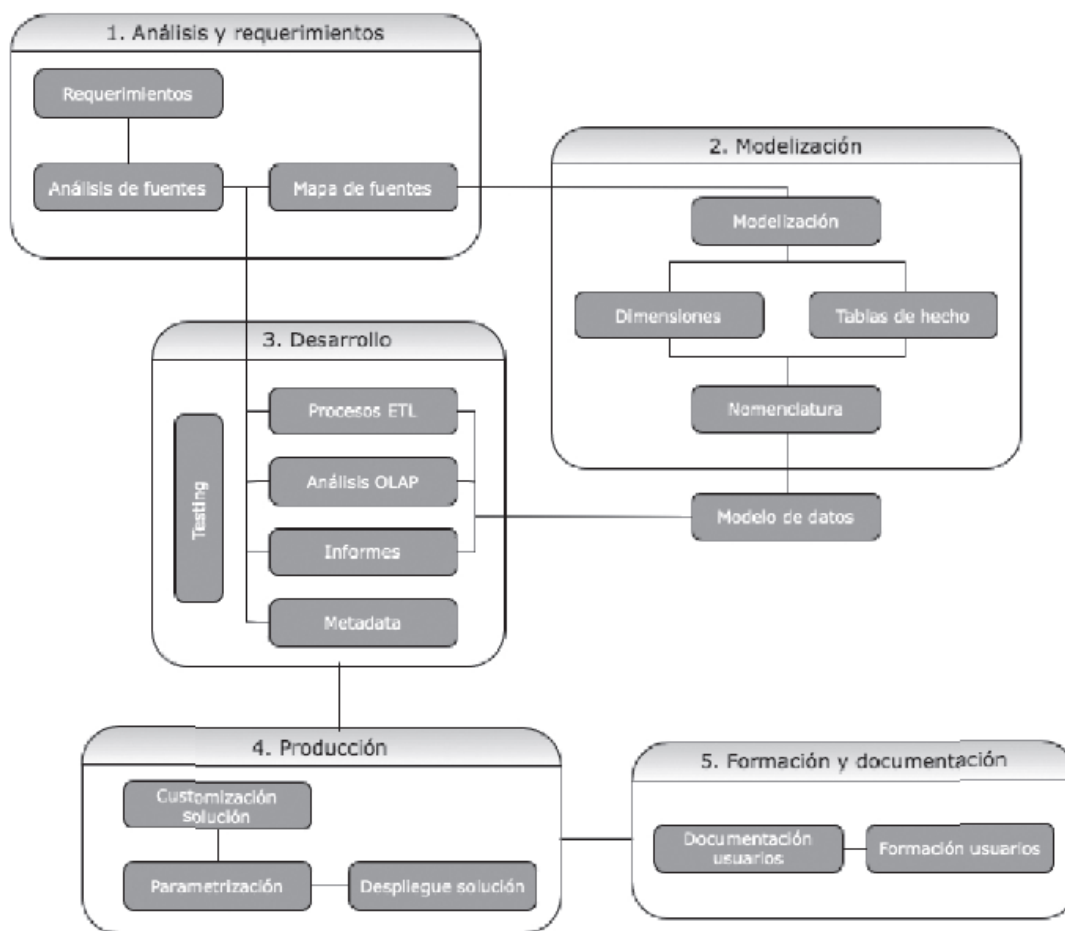


Figura 13 - Fases de un proyecto de BI
Fuente: CONESA J. & CURTO J. – 2010

4.1.1 Análisis de Requerimientos

1. Determinación de objetivos del BI y necesidades del negocio

Las empresas requieren incorporar datos de redes sociales en sus sistemas de Inteligencia de Negocio o demandan de la analítica social principalmente con el propósito de entender mejor a sus clientes, monitorizar e interpretar las interacciones entre los clientes y la difusión de nuestros productos y servicios publicado en los canales de las redes sociales.

Para responder estas cuestiones del negocio a través de una herramienta de análisis de datos de las redes sociales son las siguientes, se ha determinado los siguientes indicadores de rendimiento KPI:

En Twitter:

- Vinculación emocional hacia la empresa por los usuarios de Twitter como se han referido hacia la empresa por medio de los tweets.
 - Total de tweets positivos
 - Total de tweets negativos
 - Total de tweets neutros
- Cuáles son las menciones más usadas en relación a la empresa
 - Palabras más usadas para referirse a la empresa en los tweets y/o opiniones.
- Visibilidad y exposición, cuantos seguidores se tienen, cuales son los más influyentes y la distribución geográfica.
 - Total de seguidores
 - Cuáles son los más influyentes
 - Distribución geográfica de nuestros seguidores

En Facebook:

- Vinculación emocional por parte de los usuarios de Facebook, de qué manera se refieren hacia la empresa a través de las opiniones y publicaciones de los visitantes a la página de fans de la empresa.

- Total de opiniones positivas
- Total de opiniones negativas
- Total de opiniones neutras
- Visibilidad, exposición y alcance
 - Total de seguidores (fans)
 - Total de me gusta
- Interacción de las publicaciones
 - Publicaciones más gustadas
 - Publicaciones mas
 - Posts más compartidos
 - Posts más comentados
- Vinculación emocional a una determinada publicación
 - Total de comentarios positivos acerca del post
 - Total de comentarios negativos del post
 - Total de comentarios neutras del post

2. Recuperación de datos e información de las redes sociales

Las redes sociales permiten recuperar parte de su información con el uso de plugins, widgets o consultas a su API⁸.

Las API para las redes sociales son bibliotecas informáticas que permiten a un programa externo, través de consultas a la API, extraer datos de la red social para su posterior análisis y generación de información. Estas consultas se pueden personalizar mediante parámetros, como el idioma de los mensajes, la fecha, región de publicación, entre otros.

Las API son muy dinámicas y están en constante evolución y cambios. Son habituales las modificaciones sobre los límites de consulta, los parámetros para personalizar las búsquedas e incluso la forma de cómo deben ser consultadas, por lo que es necesaria revisarlas muy a

⁸ API - Application Programming Interface (Interfaz de Programación de Aplicaciones)

menudo. (BARTOLÍ MUÑOZ A., BRAULIO GIL N., CURTO DÍAZ J. & PUIGVERT I PELL F., 2015, pág. 32).

3. Herramientas de recuperación de datos

Para recuperar los datos de Twitter se pueden utilizar bibliotecas que interactúen con la API de Twitter⁹, alguna herramienta web o utilizar la búsqueda avanzada de Twitter.

- La interfaz web de Twitter y su búsqueda avanzada¹⁰ son una buena opción para realizar una primera exploración de forma rápida, la búsqueda avanzada permite aplicar filtros en las consultas, sin embargo, solo nos sirve para realizar una primera exploración visual, no permite automatizar la extracción y almacenamiento de datos.
- Otra opción es utilizar una herramienta comercial ya disponible que se adapte a nuestras necesidades, pero se deben tener en cuenta los siguientes inconvenientes que puede tener una herramienta externa:
 - o No está garantizada la continuidad de la herramienta.
 - o Tampoco está garantizada la actualización de la herramienta para adaptarse a las novedades en los datos que nos permite recuperar Twitter.
- Finalmente definimos la API de Twitter, la cual permite aplicar los filtros necesarios para segmentar la recuperación de datos, permite recuperar los mensajes escritos sobre un tema de forma automatizada y realizar las consultas en intervalos cortos de tiempo.

La API de Twitter ofrece una serie de parámetros para filtrar los datos a recuperar, entre los que destacan se detallan en la Tabla 7:

Nombre del parámetro	Significado
<i>Q</i> (obligatorio)	La cadena de texto que se busca
<i>Geocode</i>	Devuelve solo los tweets de los usuarios localizados en un radio <i>x</i> de las coordenadas geográficas indicadas
<i>Lang</i>	Devuelve solo los tweets en el idioma indicado
<i>Locale</i>	Indica el idioma utilizado en la cadena de texto que se busca

⁹ La API de Twitter, permite el acceso a búsquedas, acceso al motor de datos y acceso en tiempo real a los tweets públicos en Twitter.

¹⁰ Búsqueda avanzada de Twitter, <https://twitter.com/search-advanced>

Nombre del parámetro	Significado
<i>result_type</i>	Indica el tipo de resultados que se desean recibir. El valor por defecto es <i>mixed</i> y las posibilidades son: <ul style="list-style-type: none"> • <i>recent</i>: los resultados más recientes • <i>popular</i>: los resultados más populares • <i>mixed</i>: tuits populares y recientes
<i>Until</i>	Tweets generados hasta la fecha indicada
<i>Since</i>	Tweets generados desde la fecha indicada

Tabla 7 - Parámetros del Tweet

Fuente: <https://twitter.com/>

Con estos parámetros podemos filtrar los tweets por idioma, por fecha y recuperar solo aquellos mensajes ubicados en una cierta área geográfica, aunque debemos tener en cuenta que la API de Twitter a través de la búsqueda estándar (Standard Search API) permite la recuperación de los tweets publicados en los últimos 7 días. Mas información de los recursos (Tweets) se detallan en la Tabla 8 a continuación:

Información del recurso	
Formatos de respuesta	JSON
¿Requiere autenticación?	SI
¿Solicitudes limitadas?	SI
Solicitudes/15 min (Autenticación de usuario)	180
Solicitudes/15 min (Autenticación de aplicación)	450

Tabla 8 - Información de los recursos (Tweets)

Fuente: <https://twitter.com/>

Vistas las posibilidades de recuperación de información y las restricciones, se considera que la API de Twitter es la mejor opción para recuperar los mensajes escritos por los usuarios.

4. Recuperación datos a través de la API de Twitter

a) Acceso a la API de Twitter

Para tener acceso a la API de Twitter se tiene que utilizar el sistema de autenticación basada en tokens OAuth, que es un estándar abierto de autorización. A partir de la versión REST API 1.1

de Twitter se establece el uso de la autenticación OAuth como obligatoria para recuperar datos de Twitter.

El primer paso es dar de alta una aplicación en el perfil de desarrolladores de Twitter¹¹, para obtener cuatro cadenas también llamados “tokens de acceso” los cuales son necesarios para la autenticación OAuth.

b) Creación de aplicación en Twitter

Para crear una aplicación se deben seguir los siguientes a través de la interfaz (ver Figura 14) del sitio de desarrolladores de Twitter (<https://developer.twitter.com/en/apps>).

1. Hacer login en la web de Twitter
2. Ir a la página de aplicaciones de Twitter: <https://apps.twitter.com>
3. En la página de aplicaciones de Twitter seleccionamos "Create a new App".
4. En los campos “Name” y “Descripción” podemos poner los datos que queramos.
5. En el campo “Website” tenemos que poner una URL válida, aunque no es necesario que la URL exista o que sea de nuestra propiedad (P/ejemplo <http://juanperez.wordpress.com/>)
6. En el campo “Callback Url” se debe indicar una dirección URL válida, aunque tampoco es necesario que la URL exista o que sea de nuestra propiedad.
7. Leer y aceptar las condiciones de uso y seleccionar "Create your Twitter Application".
8. Seremos redirigidos a la página de datos de la App y tendremos que dirigirnos a la pestaña "Keys and Access Tokens".
9. Copiamos los datos:
 - a. Consumer Key (API Key)
 - b. Consumer Secret (API Secret)
10. Al final de esta página tenemos acceso a nuestro “Access Token” y copiamos los datos:
 - a. Access Token
 - b. Access Token Secret

¹¹ Acceso al sitio de desarrolladores de Twitter <https://developer.twitter.com/en/apps>

The screenshot shows the Twitter Application Management interface for an application named 'RecuperacionDatos'. The page is divided into several sections:

- Application Settings:** Contains fields for 'Consumer Key (API Key)', 'Consumer Secret (API Secret)', 'Access Level' (Read, write, and direct messages), 'Owner' (CondoriMaritza), and 'Owner ID' (2577025682).
- Application Actions:** Includes buttons for 'Regenerate Consumer Key and Secret' and 'Change App Permissions'.
- Your Access Token:** Contains fields for 'Access Token', 'Access Token Secret', 'Access Level' (Read and write), 'Owner' (CondoriMaritza), and 'Owner ID' (2577025682).
- Token Actions:** Includes buttons for 'Regenerate My Access Token and Token Secret' and 'Revoke Token Access'.

*Figura 14 - Claves y tokens de acceso a la API de Twitter
Fuente: <https://twitter.com/>*

5. Herramientas para la recuperación de datos

La recuperación de datos puede realizarse mediante librerías de R o Python, en nuestro caso usaremos la librería `twitterR`¹² para realizar una búsqueda de datos en la red social.

¹² Librería `twitterR`, <https://cran.r-project.org/web/packages/twitterR/README.html>

a) Recuperación de tweets con R (tweets vinculados a un tema)

```
#Cargar la librería twitterR
library("twitterR")

#Establecemos las cuatro cadenas de acceso OAuth
ConsumerKey = ". . . . . ."
ConsumerSecret = ". . . . . ."
AccessToken = ". . . . . ."
AccessTokenSecret = ". . . . . ."

#Ejecutar la autenticación de TwitterR
setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_secret)

#Realización de la búsqueda de datos especificando la palabra clave
tweets <- searchTwitter("hashtag", n=200, lang="es")

#Para impresión del total de tweets recuperados
length(tweets)

# Mostrar la estructura del tweet
str(tweet)

# Obtener el texto del tweet
tweet$getText()

# Extracción de todos los datos de los tweets recuperados, mediante un Data Frame de R
tweetsDF <- twListToDF(tweets)
```

b) Recuperación de datos de los seguidores (followers)

Recuperación de los seguidores de usuarios específicos de Twitter

```
#Cargamos la librería
library(rtweet)
library(twitterR)

#Cargar las credenciales de acceso
consumer_key <- ". . . . . ."
consumer_secret <- ". . . . . ."
access_token <- ". . . . . ."
access_secret <- ". . . . . ."

#Ejecutar la autenticación de TwitterR
setup_twitter_oauth(consumer_key, consumer_secret, access_token, access_secret)

#Obtener el usuario
usuario <- getUser("usuario")
user_follower_IDs <- usuario$getFollowers(retryOnRateLimit=180)

#Para impresión del total de seguidores recuperados
length(user_follower_IDs)

# Obtención de los datos de los seguidores en un Data Frame de R
require(data.table)
user_followers <- rbindlist(lapply(user_follower_IDs, as.data.frame))
```

6. Recuperación de datos a través de API Graph de Facebook

La API Graph es el modo principal para extraer e ingresar datos en la plataforma de Facebook. Se trata de una API basada en HTTP de nivel inferior que se puede utilizar de manera programática para consultar datos y llevar a cabo varias tareas más propias de una aplicación. (FACEBOOK, 2017).

El explorador API Graph es una aplicación que permite explorar la red social de Facebook permitiendo obtener datos sus datos: información del usuario, comentarios, fotos, etc.

a) Acceso a la API Graph de Facebook

El acceso al explorador de la API Graph requiere del inicio de sesión en Facebook y abrir el explorador mediante el link: <https://developers.facebook.com/tools/explorer/> (ver Figura 15).

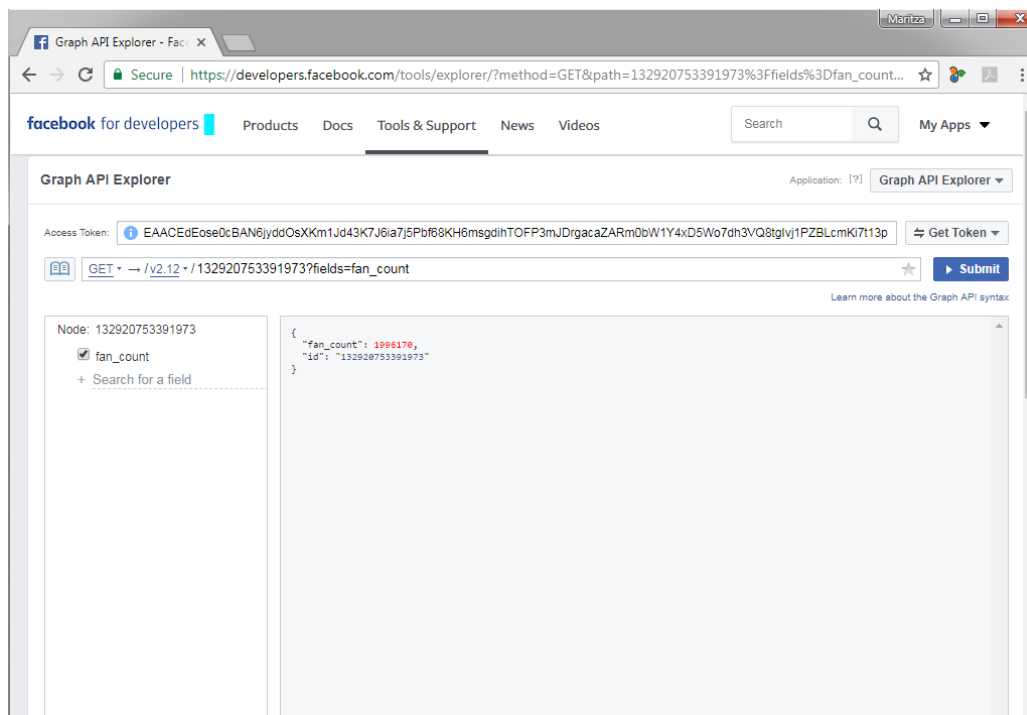
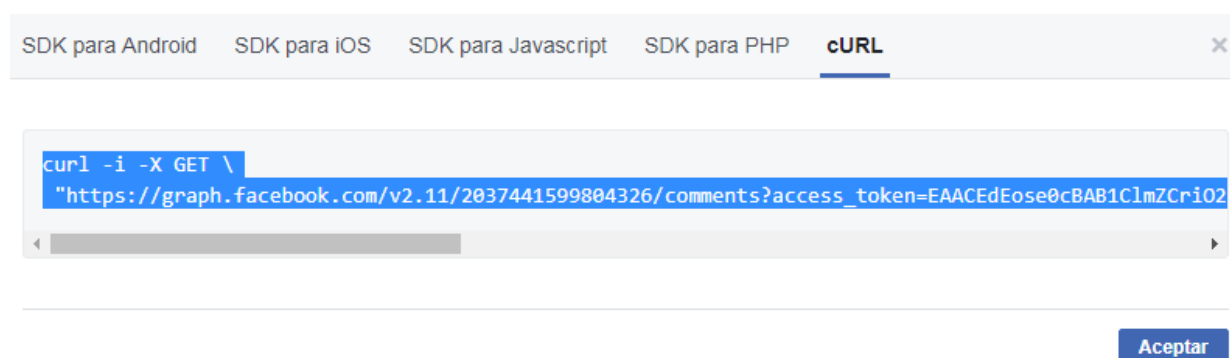


Figura 15 - Explorador de la API Graph de Facebook
Fuente: <https://developers.facebook.com/>

El acceso a la API Graph necesita el uso de tokens de acceso¹³ el cual se debe generar con la creación de una aplicación, también se puede obtener un token de acceso de usuario.

Por intermedio de esta herramienta podemos buscar y obtener varios datos en torno a nuestra red social de Facebook, tales como: número de fans, me gusta, opiniones, publicaciones, etc. Dicha información y datos son proporcionados en formato JSON.

La API Graph está basada en HTTP, por lo que funciona con cualquier lenguaje que tenga una biblioteca HTTP, como cURL y urllib (ver Figura 16).



*Figura 16 - Explorador de la API Graph de Facebook
Fuente: <https://developers.facebook.com/>*

Asimismo existen otras herramientas de acceso a la API de Facebook, tal es el caso del paquete Rfacebook¹⁴ que permite acceder a la API de Facebook a través de R.

b) Recuperación de las últimas publicaciones (posts) realizados en la página de Facebook

```
# Instalar y cargar la librería requerida
library("Rfacebook")

# Obtener información de autenticación https://developers.facebook.com/apps
appid <- " . . . . . "
appsecret <- " . . . . . "

# Autenticación de acceso a Facebook
fboauth <- fbOAuth(appid, appsecret)

# Obtener recientes posts o publicaciones de la página (fanpage)
page <- "fanpage name"
empresa_page <- getPage(page, token = fboauth)
recent_posts <- as.data.frame(empresa_page)
```

¹³ Generación de tokens de acceso : <https://developers.facebook.com/docs/facebook-login/access-tokens/>

¹⁴ Librería Rfacebook, <https://cran.r-project.org/web/packages/Rfacebook/Rfacebook.pdf>

```

1          message

Consigue tu Plan Postpago sin Límites ahora y adquiere un celular por sólo 5Bs. \nMás información aquí<u+2
7A1><U+FE0F> https://goo.gl/AE839w
2

Hoy recordamos a nuestros héroes que lucharon por
nuestro mar. #23deMarzo #DíaDelMar
3

¿Loco por mandar memes a todos tus grupos? - Si eres un cliente F&BQ con mas de 6 meses, dup
lica tus Megas con la #DobleCarga.
4

¡Feliz día internacional de la felicidad! \nEtiqueta a aquellas personas que te
ayudan a que esta fecha sea mejor.
5

¡Felicidades a todos los papás en
su día! #DíadelPadre #GraciasPapá
6

```

```

#Obtener los comentarios de un post específico
post_comments <- getPost(post = recent_posts$id[1], token = fboauth)
post_comments <- (post_comments$comments)$message

```

Los datos textuales extraídos de las redes sociales resultantes del proceso de recolección, mediante el código anteriormente descrito, los cuales son útiles para satisfacer los requerimientos, contienen elementos tales como emoticonos, hiperenlaces u otros caracteres a menudo utilizados en la red social, que no permiten la correcta ejecución del análisis, en ese sentido es necesario realizar un pre procesamiento aplicando técnicas de minería de texto (text mining) para extraer y cualificar los mismos.

7. Pre-Procesamiento de texto

El pre procesamiento de texto se realiza con el fin de homogeneizar y limpiar los mensajes de texto, entonces para analizar el texto que hay en cada mensaje como por ejemplo en los tweets es necesario considerar los siguientes puntos para su limpieza:

- Eliminar nombres de usuarios (@), por ejemplo en la red social Twitter cada usuario dispone de un alias precedido del símbolo @. Debido a que este no es un elemento gramatical puede tratárselo como nombre propio o eliminarlo del conjunto de datos.
- Eliminar hashtags (#), los hashtags son términos incluidos en los mensajes con el objetivo de etiquetar sus mensajes, para fines de análisis pueden ser borrados del

conjunto de datos o ser considerados como una palabra sostenida, en este caso únicamente el símbolo ‘#’ es borrado.

- Simplificación de enlaces, es habitual que los usuarios incorporen recursos externos en sus mensajes que hacen referencia a imágenes, videos, etc.

8. Análisis de sentimiento

En este punto se aplican tareas propias del procesamiento del lenguaje natural NPL y minería de texto TM, los lenguajes de programación R y Python cuentan con herramientas para realizar las siguientes tareas para el pre procesamiento de texto. Las tareas del pre procesamiento de texto:

a) Preprocesamiento de texto

1. Conversión de mayúsculas, eliminación de números, eliminación de direcciones URL
2. Eliminación de preposiciones y pronombres (stopwords) por que no aportan información al sentimiento o a la opinión del tweet.
3. Recortar las palabras a su raíz (stemming).
4. Construcción de una matriz de palabras, en la que cada fila corresponde a un documento de texto, cada columna a una palabra y cada celda contendrá el número de veces que aparece la palabra en el documento.
5. La matriz de palabras es el objeto central sobre el que se ejecutarán algoritmos propios de la minería de datos y machine learning, con lo cual es posible el estudio de la frecuencia de aparición de palabras así como también la clasificación del estado afectivo de los individuos.

b) Cálculo de la puntuación

Cálculo de una puntuación de confianza para cada tweet para que podamos saber lo positivo o negativo que es el mensaje publicado. En este caso vamos a utilizar un enfoque simple pero útil para definir la fórmula de calificación.

Puntuación (score) = Numero de palabras positivas – Numero de palabras negativas

- Si la puntuación es mayor (>) que 0, el mensaje tiene una opinión positiva.
- Si la puntuación es menor (<) que 0, el mensaje tiene una opinión negativa.
- Si la puntuación es igual (=) que 0, se considera que el mensaje tiene una opinión neutra.

Definición de la función `score.sentiment`¹⁵ y generación de la variable de cuantificación del estado emocional.

```
score.sentiment = function(sentences, pos.words, neg.words, .progress='none')
{
  # Parametros
  # sentences: vector de textos a calificar
  # pos.words: vector de palabras que expresan sentimiento positivo
  # neg.words: vector de palabras que expresan sentimiento negativo
  # .progress: passed to laply() para controlar la barra de progresoess bar

  # create simple array of scores with laply
  scores = laply(sentences,
  function(sentence, pos.words, neg.words)
  {
    # remove punctuation
    sentence = gsub("[[:punct:]]", "", sentence)
    # remove control characters
    sentence = gsub("[[:cntrl:]]", "", sentence)
    # remove digits?
    sentence = gsub('\\d+', '', sentence)

    # define error handling function when trying tolower
    tryTolower = function(x)
    {
      # create missing value
      y = NA
      # tryCatch error
      try_error = tryCatch(tolower(x), error=function(e) e)
      # if not an error
      if (!inherits(try_error, "error"))
      y = tolower(x)
      # result
      return(y)
    }
    # use tryTolower with sapply
    sentence = sapply(sentence, tryTolower)

    # split sentence into words with str_split (stringr package)
    word.list = str_split(sentence, "\\s+")
    words = unlist(word.list)

    # compare words to the dictionaries of positive & negative terms
    pos.matches = match(words, pos.words)
    neg.matches = match(words, neg.words)

    # get the position of the matched term or NA
    # we just want a TRUE/FALSE
    pos.matches = !is.na(pos.matches)
    neg.matches = !is.na(neg.matches)

    # final score
    score = sum(pos.matches) - sum(neg.matches)
    return(score)
  }
}
```

¹⁵ Metodo `score_sentiment`,
https://github.com/gastonstat/Mining_Twitter/blob/master/Rscripts/Sentiment_Analysis_with_Drinks.R

```

    }, pos.words, neg.words, .progress=.progress)

# data frame with scores for each sentence
scores.df = data.frame(text=sentences, score=scores)
return(scores.df)
}

```

c) Cálculo de la matriz de menciones frecuentes

Las expresiones hacia la empresa y menciones frecuentes, estos indicadores nos permiten determinar cuáles son las menciones más usadas en relación las publicaciones de productos o servicios de la empresa, es decir, palabras más usadas para referirse a un determinado producto o servicios a través:

- Opiniones y posts de productos/servicios en Facebook
- Tweets de productos/servicios en Twitter

```

#Cargar comentarios
comentarios <- read.csv("dfTweetsEJ2.csv", sep = ",")
#
library(qdap)
library(NLP)
library(tm)

#Lo colocamos en un objeto Corpus para trabajar con text Mining
corp <- Corpus(VectorSource(comentarios$text))

#PRE PROCESAMIENTO DE TEXTO
corp <- tm_map(corp,tolower)
#Le removemos los números a nuestro
corp <- tm_map(corp, removeNumbers)
#Le removemos los stopwords(la, de, el, .)
corp <- tm_map(corp,removeWords, stopwords('spanish'))
#Eliminar las propias stopwords
#Le removemos la puntuación a nuestro texto
corp <- tm_map(corp, removePunctuation)
#Eliminar espacios en blanco
corp <- tm_map(corp, stripWhitespace)
#Realizar stemming
corp <- tm_map(corp, stemDocument)

#Conteo de palabras y generación de la matriz de frecuencias
dtm <- TermDocumentMatrix(corp)
m <- as.matrix(dtm)
v <- sort(rowSums(m), decreasing = TRUE)
d <- data.frame(words = names(v), freq=v)
head(d, 30)

```

El preprocesamiento de texto se realiza con el fin de generar la matriz de menciones frecuentes, esta data es almacenada las tablas de hechos H_MENCIONES del data warehouse.

9. Guardar información de las redes en base de datos

Todos los datos recuperados y/o extraídos de las redes sociales con los mecanismos especificados en el punto anterior serán almacenados en una base de datos a partir de la cual se alimentará el modelo de datos analítico. Para fines de construcción del prototipo de la solución social BI almacenaremos los datos en una base de datos relacional, para el presente caso se usara PostgreSQL.

10. Periodicidad de recuperación de datos

Para los fines de investigación se consideró conveniente determinar la siguiente periodicidad de extracción de información.

a) Twitter:

- Los tweets vinculados a la empresa pueden ser recuperados diario.
- La extracción de seguidores (followers) se considera puede ser efectuado de manera semanal o mensual.

b) Facebook

- Proceso de carga de los posts o publicaciones de forma semanal.
- Proceso de carga de comentarios por post cada 24 a 48 horas de haberse publicado el post para medir la aceptabilidad del público al producto o servicio.
- Proceso de carga de datos generales de la comunidad (seguidores y me gusta) puede efectuarse de manera semanal o mensual.
- Opiniones y publicaciones de los visitantes a la página (fanpage) podría efectuarse de forma semanal o mensual.

11. Análisis de fuentes de datos

Para la creación de los cuadros de mando y para la obtención y validación de los indicadores que nos permitan satisfacer los requerimientos planteados, disponemos de los siguientes ficheros de datos en formato CSV extraídos de las redes sociales Twitter y Facebook.

a) Datos extraídos de Twitter:

- **Comunidad**, datos generales de la empresa en relación a los indicadores básicos de la red social, útiles para la medir la visibilidad y exposición de la empresa (ver Tabla 9).

Nombre del campo	Descripción
Followers_count	Número de seguidores
Likes_count	Número Me gusta del usuario/empresa
Tweets_count	Número de tweets realizados por el usuario/empresa
Following_count	Número de usuarios a los que la empresa sigue

*Tabla 9 - Datos comunidad Twitter
Fuente: Elaboración propia*

- **Tweets vinculados a un tema o usuario (#hashtag)**, la posibilidad de extraer los comentarios y opiniones de los usuarios en relación a un tema nos brinda una fuente valiosa para poder analizar la vinculación emocional o el sentimiento que tienen los usuarios con la empresa en la siguiente se detalla todos los datos del tweet (ver Tabla 10).

Nombre del campo	Descripción
Text	Contenido y/o texto del tweet
Favorite	Si bien este comentario ha sido favorecido por otros usuarios
favoriteCount	Número de veces que fue marcado como favorito
replyToSN	nombre de usuario que responde a
Created	Cuando se creó este tweet
Truncated	si este estado fue truncado
replyToSID	
Id	Id del estado
replyToUID	ID del usuario que fue en respuesta a
statusSource	Agente de usuario de origen para este tweet
Screenname	Nombre de pantalla del usuario que publicó este estado.
retweetCount	La cantidad de veces que se ha retuiteado este tweet.
isRetweet	Si el Tweet extraído es Retweet
Retweeted	VERDADERO si este estado ha sido retuiteado
Longitude	Un valor de longitud numérica, entre -180 y 180 inclusive. Sur es negativo, Norte es positivo
Latitude	Valor de latitud numérica, entre -180 y 180 inclusive. West es negativo, Este es positivo
Score	Sentimiento del tweet expresado en un valor positivo, negativo o cero.

*Tabla 10 - Datos del Tweet
Fuente: Elaboración propia*

- **Datos de los seguidores (followers)**, con los datos extraídos de los seguidores (ver Tabla 11) será posible satisfacer necesidades de visibilidad y exposición de la empresa, así como que usuarios podrían considerarse influenciadores.

Nombre del campo	Descripción
Description	descripción del usuario
statusesCount	Número de actualizaciones de estado que este usuario ha tenido
followersCount	número de seguidores para este usuario
favoritesCount	Número de favoritos para este usuario
friendsCount	Número de seguidores para este usuario
url	una URL asociada a este usuario
Name	Nombre del seguidor
Created	cuando este usuario fue creado
Protected	si este usuario está protegido o no
Verified	si este usuario está verificado o no
screenname	nombre de pantalla del usuario
Location	ubicación del usuario
Lang	Idioma
Id	valor de ID para este usuario
listedCount	la cantidad de veces que este usuario aparece en listas públicas
followRequestSent	si se autentica a través de OAuth, será VERDADERO si ha escuchado una solicitud de amistad para este usuario
profileImageUrl	URL de la imagen de perfil del usuario, si existe

*Tabla 11 - Datos del Follower
Fuente: Elaboración propia*

b) Datos extraídos de Facebook:

- **Comunidad (Community)**, los datos correspondientes a la comunidad asociada al canal de la red social los cuales se detallan a continuación (ver Tabla 12) permitirán medir la visibilidad y alcance que tiene la empresa.

Nombre del campo	Descripción
Likes_count	Total de Me gusta
Fan_count	Total de Seguidores

*Tabla 12 - Datos comunidad Facebook
Fuente: Elaboración propia*

- **Publicaciones (Posts)**, las publicaciones de ofertas y servicios es quizás el principal motivo de uso de la red social, saber que publicaciones son más compartidos, más gustados y más comentados puede resultar muy importante para las empresas, es así, que los siguientes datos (ver Tabla 13), nos permitirán satisfacer dichas necesidades.

Nombre del campo	Descripción
from_id	ID del usuario quien comparte el post
from_name	Nombre del usuario quien comparte el post
Message	Texto vinculado al post
created_time	Fecha de la publicación
Type	Tipo del post (video, foto, texto)
Id	ID asociado al post
Link	Link asociado al post
Story	Usuario con quien se comparte el post
likes_count	Numero de me gusta del post
comments_count	Numero de comentarios del post
shares_count	Numero de compartidos del post

*Tabla 13 - Datos publicaciones Posts Facebook
Fuente: Elaboración propia*

- **Comentarios de un determinado post**, conocer la opinión hacia una determinada promoción o anuncio es muy importante para la empresa, los siguientes datos (ver Tabla 14) nos permitirá realizar este análisis.

Nombre del campo	Descripción
post.from_id	ID del usuario quien compartió el post
post.from_name	Nombre del usuario que compartió el post
post.message	Mensaje del post
post.created_time	Fecha de la publicación
post.type	Tipo del post (video, foto, texto)
post.link	ID asociado al post
post.id	Link asociado al post
post.likes_count	Numero de me gusta del post
post.comments_count	Numero de comentarios del post
post.shares_count	Numero de compartidos del post
comments.from_id	ID del usuario que escribió el comentario
comments.from_name	Nombre del usuario realizo el comentario
comments.message	Texto del comentario
comments.created_time	Fecha del comentario
comments.likes_count	Numero de me gusta del comentario
comments.comments_count	Numero de comentario realizados al comentario
comments.id	ID del comentario

*Tabla 14 - Datos Comentarios Facebook
Fuente: Elaboración propia*

- **Opiniones (Reviews)**, mediante Facebook también podemos analizar la vinculación emocional a través de los datos de las opiniones de los usuarios (ver Tabla 15) y calificaciones realizadas por los mismos.

Nombre del campo	Descripción
Ratings Data Created Time	Fecha de emisión de la clasificación
Ratings Data Reviewer Name	Nombre del usuario que emite la opinión
Ratings Data Reviewer Id	ID usuario que emite la opinión
Ratings Data Rating	Puntuación y/o Valoración
Ratings Data Review Text	Texto de la opinión
Id	ID de la opinión

*Tabla 15 - Datos de opiniones Facebook
Fuente: Elaboración propia*

La puntuación se realiza bajo la siguiente clasificación (ver Tabla 16):

Estrellas	Descripción
5	Excelente (Excellent)
4	Muy bueno (Very Good)
3	Bueno (Good)
2	Aceptable (Fair)
1	Deficiente (Poor)

*Tabla 16 - Puntuaciones Facebook
Fuente: Elaboración propia*

- **Publicaciones de los visitantes**, es posible obtener los comentarios de los visitantes en los cuales se suele publicar las opiniones y/o consultas hacia la empresa, esta información (ver Tabla 17) también es bastante útil para determinar la vinculación emocional.

Nombre del campo	Descripción
Visitor Posts Data Created Time	Fecha de emisión de la clasificación
Visitor Posts Data Message	Texto del post o comentario
Visitor Posts Data Id	ID del post o comentario
Score	Sentimiento del post expresado en un valor positivo, negativo o cero.

*Tabla 17 - Publicaciones visitantes Facebook
Fuente: Elaboración propia*

Una vez realizada la carga inicial con todos los datos descritos se plantea: Una carga diaria de los ficheros siempre y cuando se detecten cambios en ellos, ya que la exigencia por parte de los clientes puede ser máxima y no se pudiera tomar decisiones en base a datos no actualizados.

4.1.2 Modelización (Diseño e implementación del almacén de datos)

a) Modelo de datos

Para la realización de la demo de la solución inicialmente tenemos necesidades de almacenamiento relativamente mínimas, por lo tanto no será necesario un sistema de almacenamiento que permita escalabilidad horizontal, sino que los datos del almacén se cargaran de forma periódica.

Bajo estas consideraciones se piensa que las bases de datos NoSQL documentales podría ser una buena opción de almacenamiento del modelo dimensional para el presente trabajo sin embargo este tipo de base de datos dificultaría la creación de nuevos indicadores, consultas y/o filtros de datos.

Otra opción viable sería el uso de un modelo relacional clásico debido a su gran personalización, ya que permite realizar consultas complejas para recuperar información sobre la base de distintos parámetros y permite realizar consultas ad hoc que no se hayan especificado a priori, que podrían ser útiles para validar si los indicadores son lo suficientemente predictivos o es mejor utilizar unos nuevos.

Por lo expuesto se ha determinado la utilización de una base de datos relacional debido a la madurez y a la alta integración que tienen con las suites de inteligencia de negocio.

b) Diseño del almacén de datos

A partir del análisis realizado hasta el momento se han identificado tres tablas de hechos para analizar los datos de la red social Twitter: Una para los tweets, otra para los seguidores y otra para las menciones. Para la red social Facebook se tienen las siguientes tablas de hechos: Una para los posts del usuario, otra para las opiniones del usuario, otra para los posts de la empresa y otra para la comunidad. Estas tablas de hechos estarán relacionadas en el almacén de datos mediante la fecha, que es el nexo en común de todas. El siguiente diagrama muestra el diseño conceptual del almacén de datos:

- **Diseño del modelo conceptual – Twitter**, se representa a través del siguiente esquema (ver Figura 17).

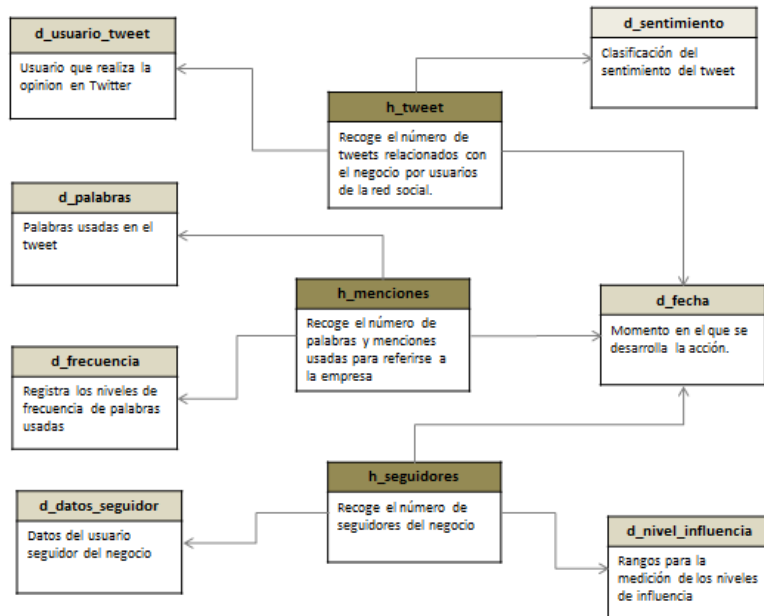


Figura 17 - Modelo conceptual (Twitter)
Fuente: Elaboracion propia

- **Diseño del modelo conceptual – Facebook**, se representa a través del siguiente esquema (ver Figura 18).

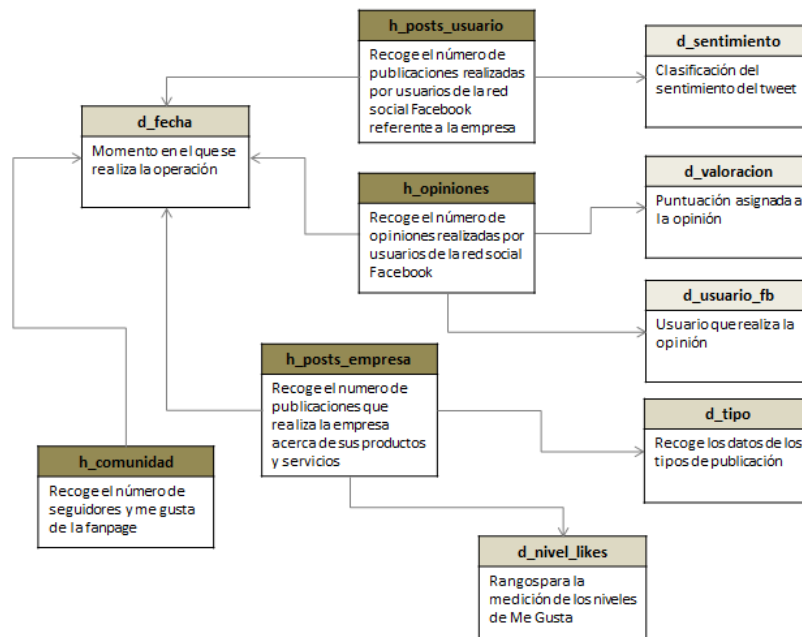


Figura 18 - Modelo conceptual (Facebook)
Fuente: Elaboracion propia

c) Diseño del modelo lógico

Una vez determinados que hechos, dimensiones, métricas y atributos existen podemos determinar el diseño lógico de la base de datos que se creará. Al utilizar el modelo relacional se deberán identificar las distintas tablas, sus atributos y las claves foráneas que representaran las relaciones entre los datos.

Tabla H_TWEET

	h_tweet	Tipo
(PK)	Id	Clave primaria
(FK)	id_fecha, id_usuario, id_sentimiento	Claves foráneas
	numero_tweets	Métrica

Tabla H_SEGUIDORES

	h_seguidores	Tipo
(PK)	Id	Clave primaria
(FK)	id_fecha, id_seguidor, id_nivel_influencia	Claves foráneas
	numero_seguidores, numero_seguidores_seguidor, numero_tweets_seguidor	Métricas

Tabla H_MENCIONES

	h_menciones	Tipo
(PK)	Id	Clave primaria
(FK)	id_fecha, id_nivel_frecuencia, id_palabra, id_tipo_palabra	Claves foráneas
	numero_palabras y frecuencia_palabra	Métricas

Tabla H_POSTS_USUARIO

	h_posts_usuario	Tipo
(PK)	Id	Clave primaria
(FK)	id_fecha, id_sentimiento	Claves foráneas
	numero_posts_usuario	Métricas

Tabla OPINIONES

	h_opiniones	Tipo
(PK)	Id	Clave primaria
(FK)	id_fecha, id_usuario, id_calificacion	Claves foráneas
	numero_opiniones	Métricas

Tabla H_POSTS_EMPRESA

	h_posts_empresa	Tipo
(PK)	Id	Clave primaria
(FK)	id_fecha, id_usuario, id_nivel_likes, id_tipo	Claves foráneas
	numero_posts_empresa, numero_likes_post, numero_compartidos_post, numero_comentarios_post	Métricas

Las dimensiones descritas en el modelo conceptual tienen los siguientes atributos:

Dimensiones	Atributos
d_sentimiento	valor, descripcion
d_fecha	anio, mes, dia, semana_anio, desc_mes, desc_dia
d_usuario_tweet	Nombre
d_seguidor	nombre, usuario
d_nivel_influencia	Descripcion
d_nivel_frecuencia	Descripcion
d_tipo_palabra	Descripcion
d_palabras	Palabra
d_usuario_fb	Nombre
d_calificacion	valor , descripcion
d_tipo	tipo_publicacion
d_nivel_likes	Descripcion

d) Diseño físico

Para implementar el almacén de datos (data warehouse) se ha decidido utilizar el gestor de base de datos PostgreSQL¹⁶ en consideración a las siguientes ventajas y características:

- Sistema de base de datos relacional orientado a objetos de distribución libre.
- Ha ganado sólida reputación en fiabilidad, integridad y corrección de datos. Es multiplataforma, se puede operar en los principales sistemas operativos, incluidos Linux, UNIX (AIX, BSD, HP-UX, macOS, Solaris) y Windows.
- Es totalmente compatible con ACID, tiene soporte completo para claves externas.
- Buena escalabilidad ya que es capaz de ajustarse al número de CPU y a la cantidad de memoria disponible de forma óptima, soportando una mayor cantidad de peticiones simultáneas a la base de datos de forma correcta.

Primero se crearán las tablas de dimensión y por último las tablas de hechos.

1) Dimensión usuario: d_usuario_tweet

Campo	Tipo de dato	Descripción
id_usuario	integer	clave primaria (PK)
Nombre	character varying(20)	guardará el nombre del usuario

2) Dimensión fecha: d_fecha

Campo	Tipo de dato	Descripción
id_fecha	integer	clave primaria (PK)
Fecha	Date	fecha complete
Anio	integer	año de la fecha
Mes	integer	mes de la fecha
desc_mes	character varying(20)	mes de la fecha, literal
semana_anio	integer	semana del año
desc_dia	character varying(20)	día de la fecha, literal

3) Dimensión valoración del sentimiento: d_sentimiento

¹⁶ PostgreSQL es un potente sistema de base de datos relacional de objetos de código abierto, <https://www.postgresql.org/>

Campo	Tipo de dato	Descripción
id_sentimiento	integer	clave primaria (PK)
Valor	integer	puntuación del tweet
Descripcion	char(10)	descripción del tweet u opinión

4) Tabla de hechos de Twitter: h_tweet

Campo	Tipo de dato	Descripción
id_tweet	integer	clave primaria (PK)
id_fecha, id_usuario, id_sentimiento, id_hashtag	integer	claves foraneas (FK)
numero_tweets	integer	Metrica

5) Dimensión de datos seguidores: d_seguidor

Campo	Tipo de dato	Descripción
id_seguidor	integer	clave primaria (PK)
Nombre	character varying(30)	
Usuario	character varying(30)	

6) Dimensión de los niveles de influencia de seguidores: d_nivel_influencia

Campo	Tipo de dato	Descripción
id_nivel_influencia	integer	clave primaria (PK)
Descripcion	character varying(30)	

7) Tabla de hechos de seguidores Twitter: h_seguidores

Campo	Tipo de dato	Descripción
id_seguidores	integer	clave primaria (PK)
id_fecha, id_seguidor, id_nivel_influencia	integer	claves foraneas (FK)
numero_seguidores, numero_seguidores_seguidor	integer	Metricas

8) Dimensión de los niveles de frecuencia: d_nivel_frecuencia

Campo	Tipo de dato	Descripción
id_nivel_frecuencia	Integer	clave primaria (PK)
Descripcion	character varying(30)	

9) Dimensión de los tipos de palabra: d_tipo_palabra

Campo	Tipo de dato	Descripción
id_tipo_palabra	Integer	clave primaria (PK)
Descripcion	character varying(30)	

10) Dimensión de los tipos de palabra: d_palabras

Campo	Tipo de dato	Descripción
id_palabra	Integer	clave primaria (PK)
Palabra	character varying(20)	

11) Tabla de hechos de menciones: h_menciones

Campo	Tipo de dato	Descripción
Id	Integer	clave primaria (PK)
id_fecha, id_nivel_frecuencia, id_palabra, id_tipo_palabra	Integer	claves foraneas (FK)
numero_palabras, frecuencia_palabra	Integer	Metricas

12) Tabla de hechos: h_posts_usuario

Campo	Tipo de dato	Descripción
Id	Integer	clave primaria (PK)
id_fecha, id_sentimiento	Integer	claves foraneas (FK)
numero_posts_usuario	Integer	Metricas

13) Dimensión de los usuarios de facebook: d_usuario_fb

Campo	Tipo de dato	Descripción
id_usuario	Integer	clave primaria (PK)

Nombre	character varying(20)	
--------	-----------------------	--

14) Dimensión de las calificaciones y/o puntuaciones: d_calificacion

Campo	Tipo de dato	Descripción
id_calificacion	Integer	clave primaria (PK)
Valor	Integer	
Descripcion	character varying(20)	

15) Tabla de hechos: h_opiniones

Campo	Tipo de dato	Descripción
Id	Integer	clave primaria (PK)
id_fecha, id_usuario, id_calificacion	Integer	claves foraneas (FK)
numero_opiniones	Integer	Metricas

16) Dimensión de los tipos de publicaciones: d_tipo

Campo	Tipo de dato	Descripción
id_tipo	Integer	clave primaria (PK)
tipo_publicacion	character varying(20)	

17) Dimensión de los niveles de me gusta: d_nivel_likes

Campo	Tipo de dato	Descripción
id_nivel_likes	Integer	clave primaria (PK)
Descripcion	character varying(20)	

18) Tabla de hechos de las publicaciones de las empresas: h_posts_empresa

Campo	Tipo de dato	Descripción
Id	Integer	clave primaria (PK)
id_fecha, id_nivel_likes, id_tipo	Integer	claves foraneas (FK)
numero_posts_empresa, numero_likes_post, numero_compartidos_post, numero_comentarios_post	Integer	Metricas

e) Otras consideraciones generales para el diseño del modelo de datos

- Se han considerado claves de tipo entero e independiente de las fuentes de origen.
- Las métricas son numéricas aditivas y semiaditivas.

f) Necesidades de almacenamiento

Una de las principales necesidades para la presente solución son los datos para el análisis de sentimiento y vinculación emocional a partir de los mensajes escritos por los usuarios de Twitter se utilizan datos de los tweets sobre la empresa, así como los datos comentarios emitidos de Facebook.

El número de campos que se almacenará para cada tweet en el almacén de datos será reducido: el nombre del usuario, el enlace al perfil del autor del tweet y la fecha en la que fue escrito. Por tanto el tamaño requerido por cada tweet será bajo, así como también los comentarios de Facebook. Dependiendo de la periodicidad con la que se extraerá la información se considera que no será demasiado elevado en el contexto de los almacenes de datos.

En ese sentido, se establece que las necesidades de almacenamiento son mínimas, no es necesario un sistema de almacenamiento que permita escalabilidad horizontal, los datos del almacén de datos se cargarán de forma periódica.

El modelo documental es una buena opción ya se tienen pocos datos para calcular los indicadores (Mensajes, tweets, número de mensajes escritos, número de seguidores, número de retweets y número de favoritos en los mensajes) son pocos y reutilizables entre indicadores. Eso permitiría pre calcular estos datos en distintos agregados durante los procesos ETL y guardarlos como agregados, lo que permite calcular los indicadores de forma rápida y sencilla.

En este caso se utilizará una base de datos documental para que guarde los tweets de un día o de los últimos días y los vuelque a un almacén de datos.

4.1.3 Desarrollo

a) Diseño e implementación de la carga de datos en el almacén de datos

En el contexto de la inteligencia de negocios, las herramientas ETL son la opción usual para alimentar el almacén de datos también llamado data warehouse.

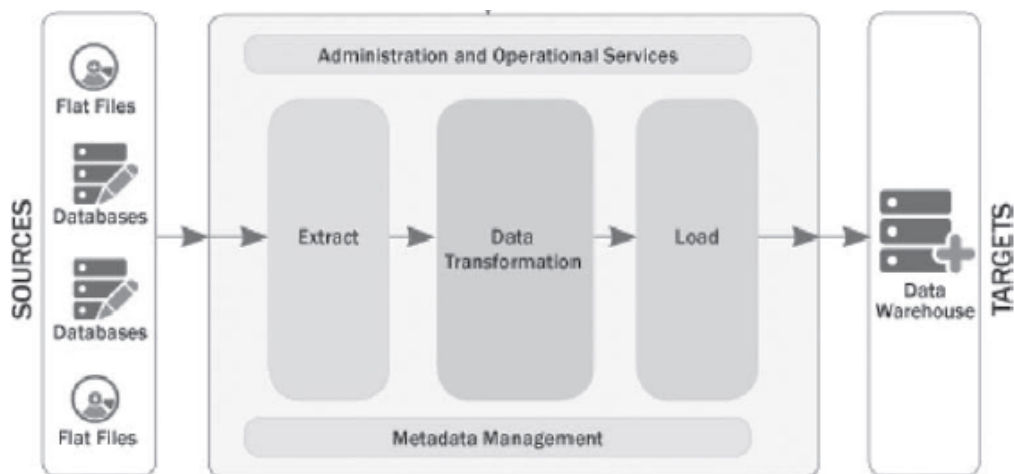


Figura 19 - Procesamiento Extracción, Transformación y Carga
Fuente: CONESA J. & CURTO J. – 2010

Una vez creado el almacén de datos (data warehouse) en este apartado nos centramos en la creación de los procesos ETL para cargar sus datos. Estos procesos consisten en la extracción, transformación y carga de los datos (ver Figura 19).

Conociendo las necesidades informacionales, se ha preparado una colección de ficheros CSV que contienen información que permitirá facilitar la carga de algunas dimensiones de nuestro data warehouse. La situación de partida está formada por los siguientes ficheros con los que se procederá a una carga inicial del almacén de datos.

Información de Twitter:

- dfTweets.csv: contiene la información y los datos de cada tweet.
- dfFollowers.csv: contiene la información de los seguidores.
- dfFrecPalabras.csv:

Información de Facebook:

- dfRatingsReviews.csv: contiene la información de las opiniones emitidas por los usuarios con la respectiva calificación.

- dfVisitorsPosts.csv: contiene la información pre procesada de los mensajes publicados por los usuarios en relación a la empresa.
- dfRecentPosts.csv: contiene la información de los post publicados por la empresa.
- dfCommunity.csv: contiene la información general en cuanto a likes y followers del fanpage de la empresa.

b) Identificación de procesos ETL necesarios

En nuestro caso tenemos se han identificado los siguientes procesos de carga diferentes:

1. Carga inicial del conjunto de datos e información de Twitter y Facebook.
2. Proceso de carga diario del conjunto de tweets.
3. Proceso de carga diario de las menciones más frecuentes en Twitter.
4. Proceso de carga en periodos de tiempo más elevados de los seguidores en Twitter.
5. Proceso de carga diario de las opiniones de los usuarios en Facebook.
6. Proceso de carga diario de las publicaciones de los usuarios en Facebook.
7. Proceso de carga diario de las publicaciones de la empresa en Facebook.
8. Proceso de carga en periodos de tiempo más elevados de la comunidad de la página de Facebook.

Los valores que se insertarán en las tablas de dimensiones son valores no fijos que se extraerán, transformarán y cargarán mediante procesos ETL. Los valores fijos serán calculados y establecidos mediante procesos ETL. Asimismo, las tablas de hechos también se cargarán mediante procesos ETL, implementaremos estos pasos de extracción, transformación y carga de datos con Pentaho Data Integration¹⁷.

Para respetar la integridad referencial del diseño físico del almacén de datos, las dimensiones se cargarán antes que las tablas de hecho.

c) Diseño de procesos ETL con Pentaho Data Integration – PDI

¹⁷ Pentaho Data Integration es una herramienta open source que ofrece buenas capacidades de Extracción, Transformación y Carga (ETL), <https://community.hds.com/docs>

En este apartado vamos a describir la funcionalidad a alto nivel de cada uno de los procesos ETL identificados.

1. Carga inicial de la dimensión *d_fecha*

El proceso ETL para la carga de datos en la dimensión *d_fecha* consistirá en los siguientes pasos básicos, el paso para el cálculo de atributos derivados consiste en que a partir de cada fecha se calculará (y cargará en la base de datos) el mes, el año, la semana, el día de la misma. El proceso ETL consistirá en:

- Extraer los valores del fichero Excel.
- Omitir campos innecesarios.
- Renombrar campos necesarios.
- Cálculo de atributos derivados de fecha.
- Cargar los datos en la tabla correspondiente.

2. Carga inicial de la dimensión *d_usuario_tweet*

El proceso ETL para la carga de datos en la dimensión *usuario Twitter* consistirá en:

- Extraer los valores del fichero CSV.
- Omitir campos innecesarios.
- Renombrar campos necesarios.
- Cargar los datos en la tabla correspondiente.

3. Carga inicial de la dimensión *d_sentimiento*

El proceso ETL para la carga de datos en la dimensión *d_sentimiento* consistirá en:

- Extraer los valores del fichero CSV.
- Omitir campos innecesarios.
- Renombrar campos necesarios.
- Establecer la descripción del sentimiento.
- Cargar los datos en la tabla correspondiente.

4. Carga inicial de la tabla de hechos h_tweet

Todas las tablas de hechos tendrán las mismas subtareas para la carga de datos, las cuales representan un proceso ETL realizada en Pentaho PDI (ver Figura 20):

1. Extraer valores.
2. Omitir campos innecesarios.
3. Establecer el contador de tweets.
4. Recuperar id para cada una de las dimensiones.
5. Cargar tabla de hecho.

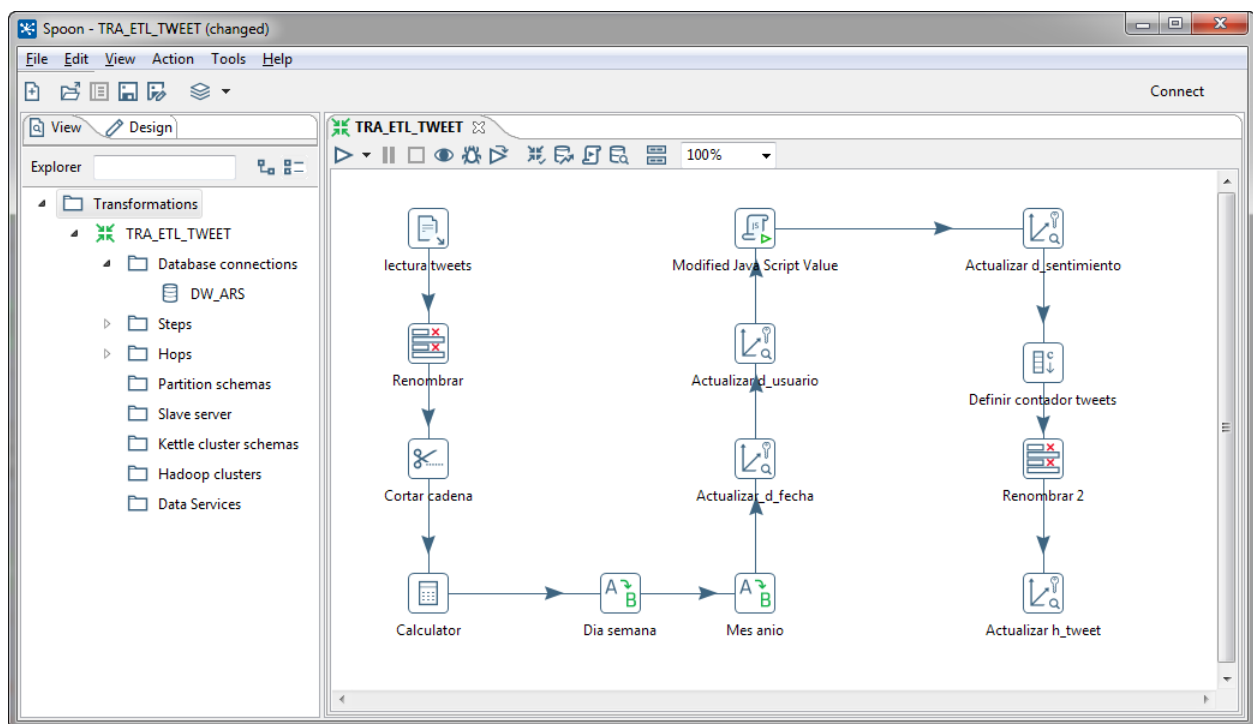


Figura 20 - Proceso ETL – Tabla de hechos H_TWEET

Fuente: Elaboración propia

d) Cargas de datos incrementales

Además de la carga de datos inicial, debemos definir los procesos de carga de datos incrementales de las diferentes fuentes de datos.

Para el fichero de datos de Twitter, se cargarán datos de forma incremental en la dimensión usuario, fecha, sentimiento y en la tabla de hechos h_tweet.

El proceso incremental consistirá en recuperar los datos de Twitter tal y como se ha explicado en la primera parte y se realizarán los mismos procesos descritos en la carga de datos iniciales:

- Extraer valores.
- Omitir campos innecesarios.
- Cálculo de atributos derivados (solo para la dimensión fecha y sentimiento).
- Recuperar id para cada una de las dimensiones (solo en el caso de la tabla de hechos h_tweet).
- Cargar tabla de dimensión o hechos.

Para realizar los procesos de carga incremental de los datos de Facebook, se deberá recuperar los datos tal como se ha explicado en la primera parte de este capítulo, posteriormente se procederá a la inserción de datos en las tablas de dimensiones y en las tablas de hechos tal y como se ha descrito anteriormente.

e) Diseño de informes y cuadros de mando (dashboards)

El punto de entrada tradicional para una herramienta de inteligencia de negocio en el contexto de una organización es la necesidad de informes operacionales y cuadros de mando.

Dado que el BI debe servir para la toma de decisiones, la información se tiene que presentar de forma resumida, a ser posible gráficamente, y mostrando solo aquellos elementos relevantes que necesitan interpretación. Para ello, se ha trabajado con herramientas de BI que sinteticen la información y muestren cuadros de mando efectivos para la toma de decisiones. En presente trabajo nos decantamos por Microsoft Power BI.

A continuación, se presentan los diferentes cuadros de mando diseñados para satisfacer los requerimientos planteados.

- Cuadros de mando para la red social Facebook

Las siguientes figuras (ver Figura 21 y Figura 22) representan la vinculación emocional en la red social Facebook.

f DASHBOARD FACEBOOK

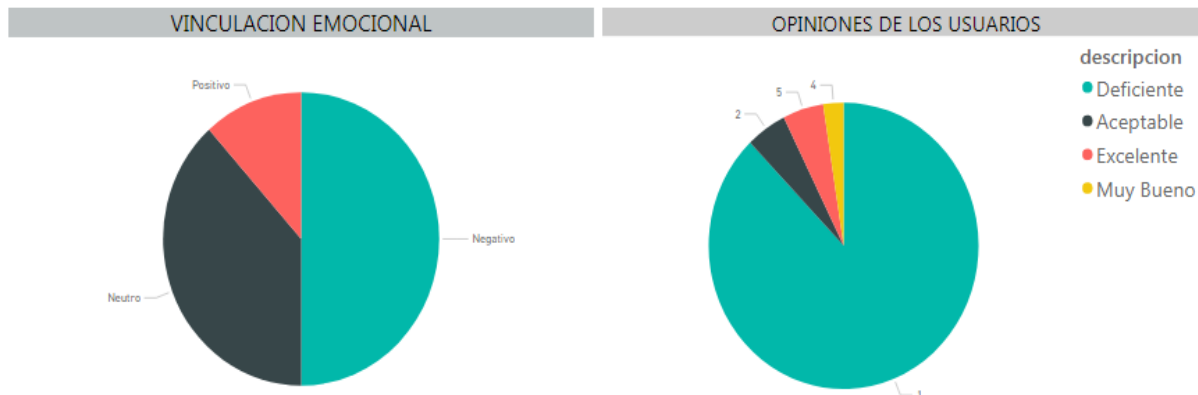


Figura 21 - Dashboard Facebook I – Vinculación Emocional
Fuente: Elaboración propia

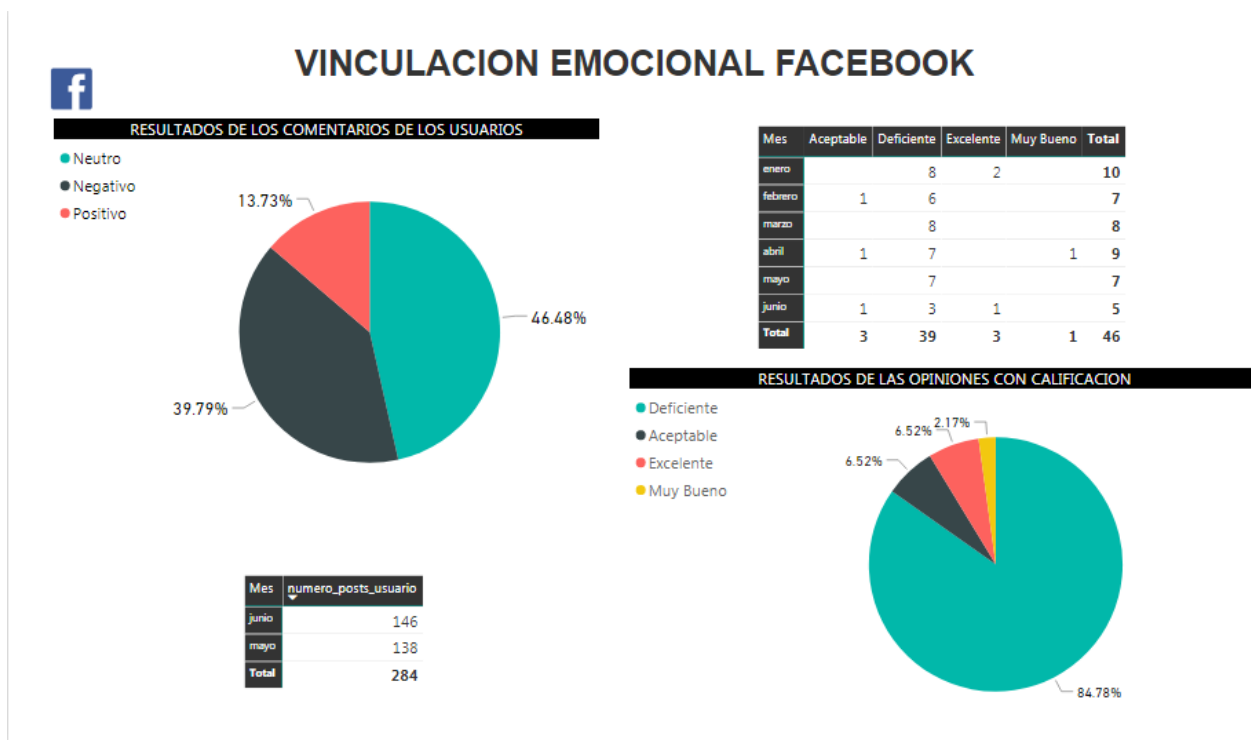


Figura 22 - Dashboard Facebook II – Vinculación Emocional
Fuente: Elaboración propia

La vinculación emocional, es la medida en la cual la gente interactúa con la empresa y su contenido. La vinculación emocional muestra las personas que están interesadas en lo que la empresa ofrece y están dispuestas a participar.

Medir este elemento es importante para ver cuánto y cómo de frecuente el público participa e interactúa con el contenido de la empresa. Para medir esta variable en Facebook hemos utilizado los siguientes indicadores:

- Menciones (positivas, negativas o neutras).
- Opiniones y Valoraciones hacia la empresa (Aceptable, Deficiente, Excelente, Muy Bueno)

De acuerdo a los resultados obtenidos en los cuadros de mando: Dashboard Facebook I – Vinculación Emocional y Dashboard Facebook II – Vinculación Emocional, la empresa debe **tomar de decisiones** específicas y determinar qué acciones y estrategias implementar para mejorar su imagen institucional de la empresa lo cual le permita obtener mejores valoraciones de los clientes.

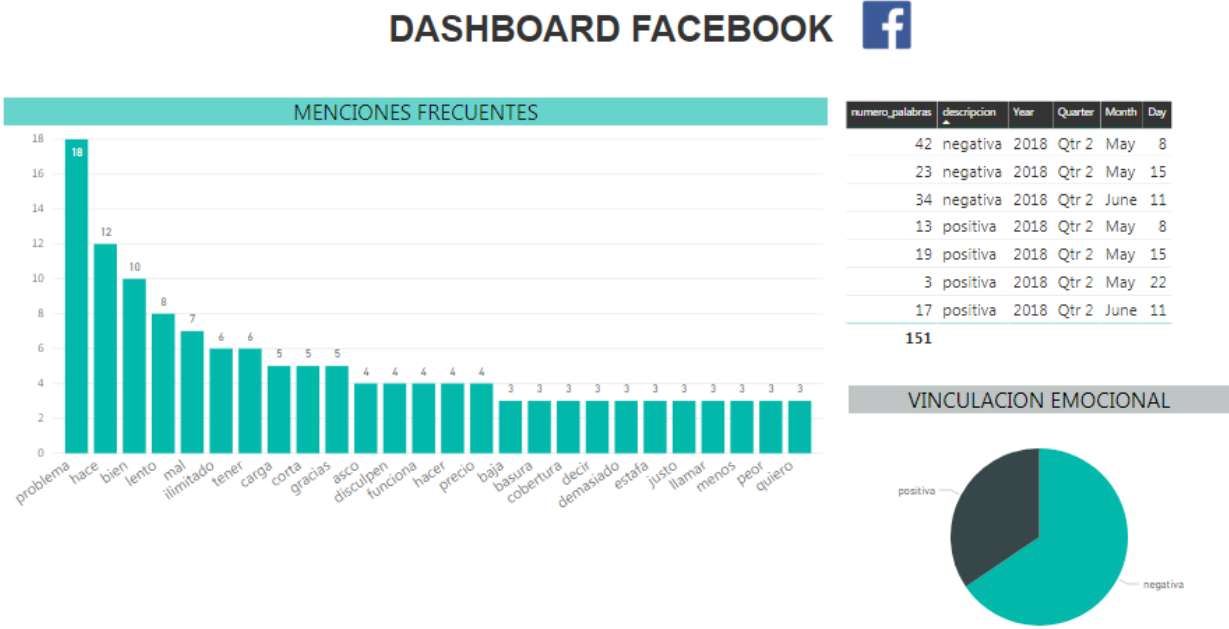


Figura 23 - Dashboard Facebook Menciones Frecuentes
Fuente: Elaboración propia

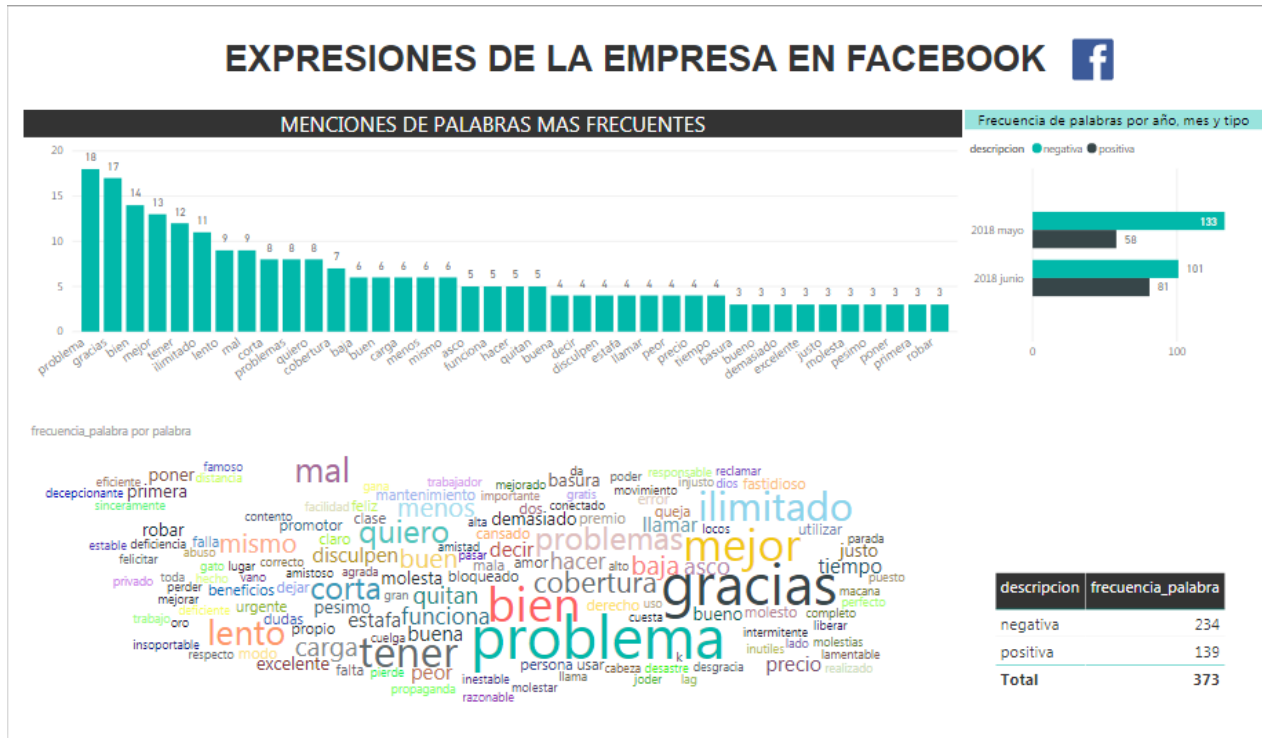


Figura 24 - Dashboard Facebook Menciones Frecuentes
Fuente: Elaboración propia

Las expresiones hacia la empresa y menciones frecuentes los cuales se exponen en los cuadros de mando para Facebook (ver Figura 23 y Figura 24) y para la red social Twitter (ver Figura 27), estos indicadores nos permiten determinar cuáles son las menciones más usadas en relación las publicaciones de productos o servicios de la empresa, es decir, palabras más usadas para referirse a un determinado producto o servicios a través:

- Opiniones y posts de productos/servicios en Facebook

Más concretamente se puede determinar la aceptación o no de un producto o servicio, en ese sentido la **toma de decisiones** se enfocaría en:

- Analizar las causas de aceptabilidad y/o rechazo del servicio, y en ese sentido tomar acciones necesarias para mejorar la atención al cliente.

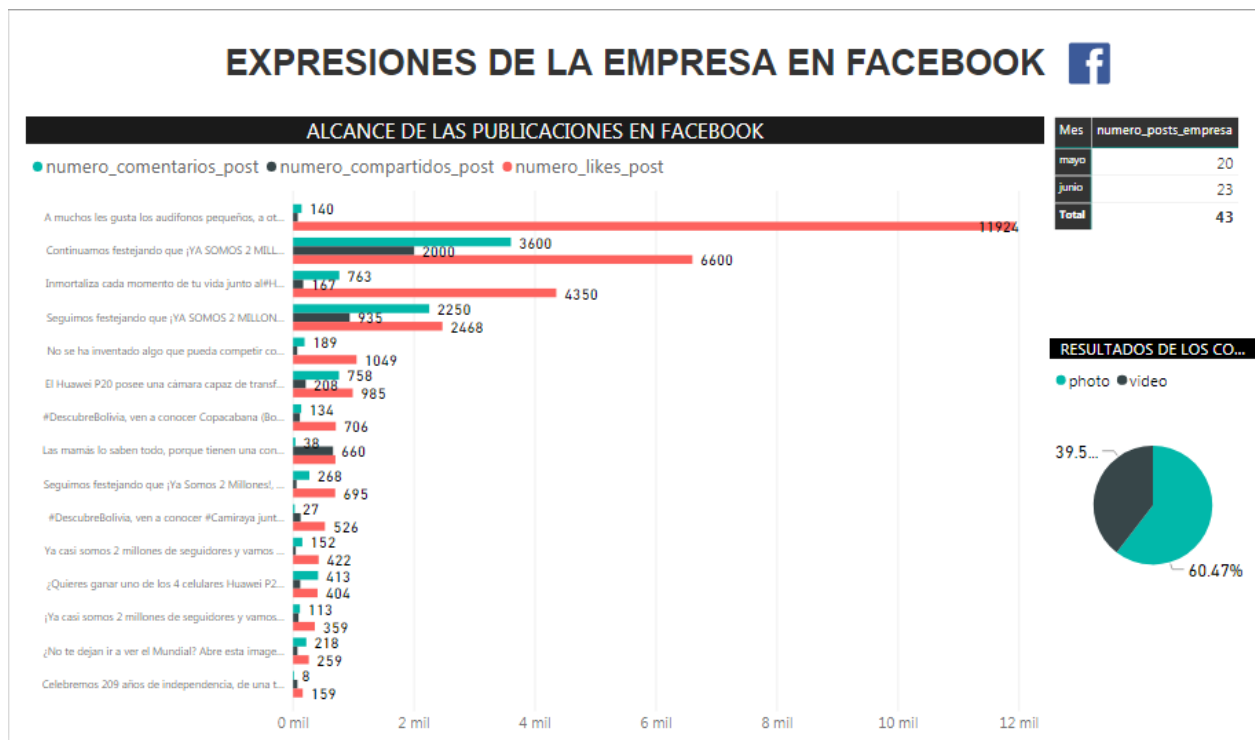


Figura 25 - Dashboard Facebook Opiniones
Fuente: Elaboración propia

Alcance de las publicaciones (posts) en Facebook, mediante este cuadro de mando (ver Figura 25) podemos ver el alcance o llegada que han tenido los posts de los diferentes productos y/o servicios y otros asuntos de la empresa difundidos a través de Facebook, más concretamente se ven con claridad los siguientes resultados:

- Productos o servicios más y menos gustados
- Productos o servicios más y menos comentados
- Productos o servicios más y menos compartidos

Es así que la **toma de decisiones** podría enfocarse entre otros en los siguientes aspectos:

- Aumentar stocks de productos o capacidad de servicios en caso de los más aceptados y demandados.
- Retirar productos y promociones de servicios que no tienes buena aceptación.
- Mejorar las estrategias de marketing de los productos y servicios para aumentar su alcance y difusión.

Cuadros de mando para Twitter

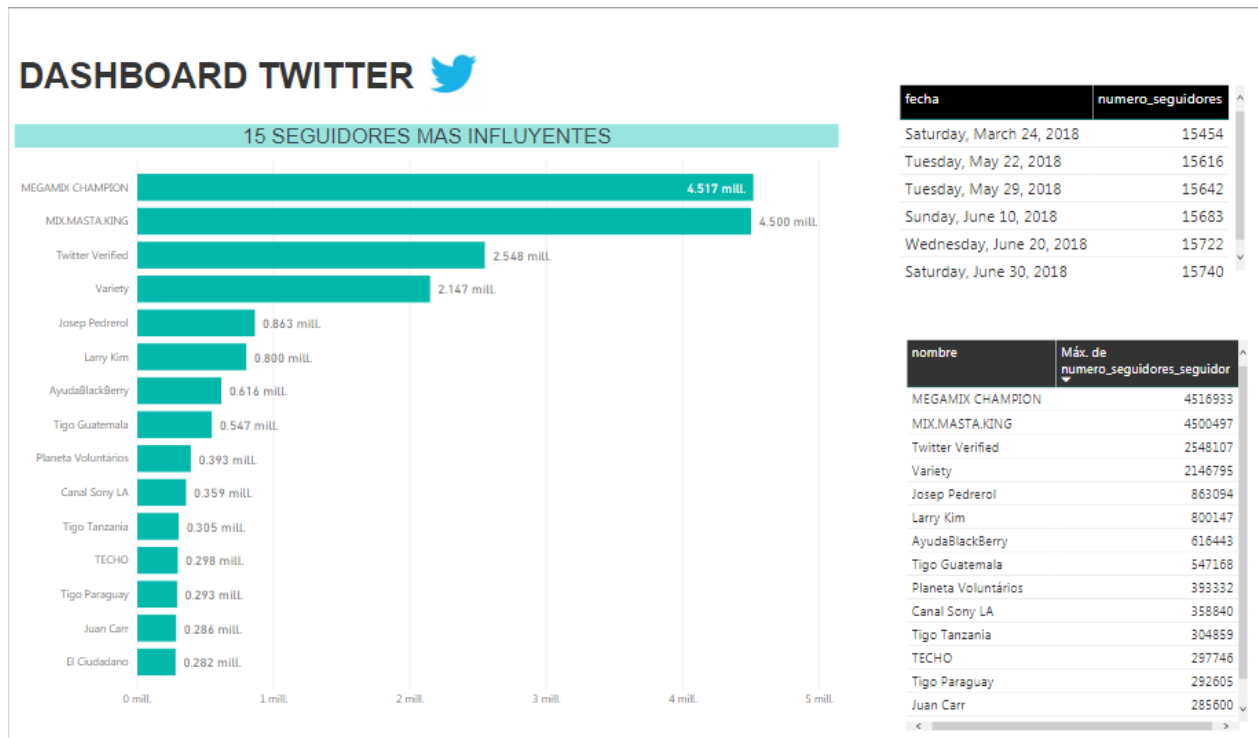


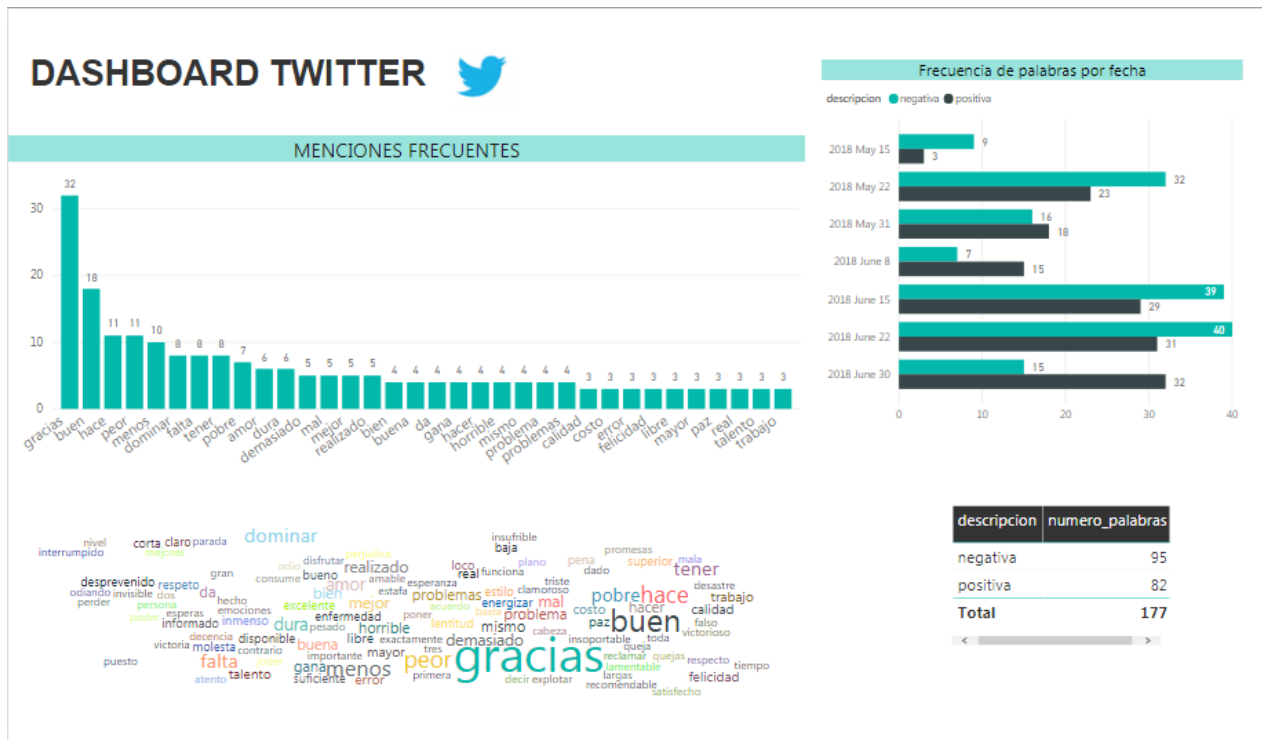
Figura 26 - Dashboard Twitter Influencers
Fuente: Elaboración propia

La influencia en redes sociales es la capacidad de un usuario de modificar el comportamiento o la toma de decisiones de sus seguidores, el cuadro de mando precedente (ver Figura 26) muestra a los seguidores más influyentes en la red social Twitter.

Un influenciador, también denominado líder de opinión, es una persona de la cual la audiencia se fía o confía, así que, cuando un influenciador comparte opiniones, pensamientos, reflexiones o ideas, la audiencia está más dispuesta a adoptar su mensaje o estar de acuerdo con él.

Conociendo que seguidores son los más influyentes para la empresa, la **toma de decisiones** se centra en determinar acciones con influencers para nuestras campañas de marketing, tales como:

- Publicaciones del influencer en nuestra fanpage de Facebook.
- Publicaciones del influencer en nuestra página de Twitter.
- Ofrecer regalos y otros.



*Figura 27 - Dashboard Twitter Menciones Frecuentes
Fuente: Elaboración propia*



*Figura 28 - Dashboard Visibilidad y Exposición en Redes Sociales
Fuente: Elaboración propia*

La visibilidad y exposición en redes sociales (ver Figura 28) se refiere a cuantos seguidores se tienen en la red social, nos permite determinar el tamaño y alcance que tiene la empresa de acuerdo a los siguientes indicadores:

- Total de seguidores (fans) en Facebook
- Total de me gusta en Facebook
- Total de seguidores o Followers en Twitter

A mayor alcance mayor es la capacidad de difusión de nuestros productos y servicios, en ese sentido la **toma de decisiones** en relación a este aspecto se podría dirigir en implementar estrategias para mejorar e incrementar la visibilidad y el alcance de la empresa en Twitter y Facebook lo cual le permitirá a la empresa sacar mayor ventaja al uso de los canales de redes sociales.

4.2 Validación de resultados

4.2.1 Estado de los objetivos y resultados alcanzados

En la Tabla 18 se detallan las acciones realizadas para el logro de los objetivos planteados:

Objetivo	Tareas realizadas
<p>Objetivo General: Desarrollar un modelo de Inteligencia de Negocios para el soporte a la toma de decisiones en base a información completa y precisa procedente de redes sociales de los consumidores y su valoración hacia los productos y servicios de la empresa.</p>	<p>Para el desarrollo del Modelo de Inteligencia de Negocios para analizar el posicionamiento de productos y servicios en Redes Sociales, se ha seguido las tres primeras fases definidas en la metodología base para el desarrollo de soluciones BI.</p> <p style="padding-left: 40px;">Fase 1: Análisis de Requerimientos (Necesidades del negocio, Recuperación de datos de Facebook y Twitter, Pre-Procesamiento de texto, Analítica Social).</p> <p style="padding-left: 40px;">Fase 2: Modelización (Diseño e implementación del almacén de datos).</p>

Objetivo	Tareas realizadas
	<p>Fase 3: Desarrollo (Diseño e implementación de la carga de datos en el almacén de datos mediante procesos ETL y Diseño de informes y cuadros de mando - dashboards para Facebook y Twitter)</p> <p>Como resultado se ha logrado el modelo de Inteligencia de Negocios que extrae, integra, analiza y explota datos e información de los consumidores proveniente de las redes sociales Facebook y Twitter, que genera el conocimiento para el soporte a la toma de decisiones de la empresa.</p> <p>Referencia: Capitulo IV-Marco Practico, Paginas 37-75</p>
<p>Objetivo específico 1</p> <p>Realizar el diseño e implementación del almacén de datos para el almacenamiento de la información proveniente de las redes sociales a través del previo análisis de requerimientos y determinación de tablas de hecho y dimensiones necesarias.</p>	<p>En la Fase 1: Análisis de Requerimientos, se ha analizado y explorado toda la información de las redes sociales y en función a ello en la Fase 2: Modelización, se ha procedido al diseño e implementación de almacén de datos de soporte al modelo de BI:</p> <ul style="list-style-type: none"> - Modelo de datos - Diseño del almacén de datos - Diseño del modelo conceptual (Twitter y Facebook) - Diseño del modelo lógico - Diseño de físico - Determinación de necesidades de almacenamiento <p>Referencia: Capitulo IV-Marco Practico, Modelización (Diseño e implementación del almacén de datos), Paginas 54-63.</p>
<p>Objetivo específico 2</p> <p>Desarrollar los procesos de extracción, transformación y carga de</p>	<p>Se ha procedido al desarrollo de los procesos de extracción, transformación y carga de datos de las redes sociales al modelo de inteligencia o data warehouse</p>

Objetivo	Tareas realizadas
<p>datos de las redes sociales al repositorio del modelo de inteligencia de negocios a través de herramientas y técnicas adecuadas.</p>	<p>diseñado para tal efecto, mediante los siguientes procesos de la fase 3:</p> <p>Fase 3: Desarrollo</p> <ol style="list-style-type: none"> 1. Diseño e implementación de la carga de datos en el almacén de datos. 2. Identificación de procesos ETL necesarios 3. Diseño de procesos ETL con Pentaho Data Integration – PDI. 4. Cargas de datos incrementales al almacén de datos. <p>Referencia: Capitulo IV-Marco Practico, 4.1.3 Desarrollo, Páginas 63-68.</p>
<p>Objetivo específico 3</p> <p>Explorar la información generada por el canal de las redes sociales Facebook y Twitter a través de técnicas y algoritmos apropiados para determinar las preferencias de los consumidores.</p>	<p>En la Fase 1: Análisis de Requerimientos, se ha procedido a la recuperación de datos e información de las redes sociales Twitter y Facebook, posteriormente se ha procedido a la aplicación de técnicas para el pre-procesamiento y del algoritmo de minería de texto (basado en tareas propias del procesamiento del lenguaje natural NPL) lo cual ha permitido determinar el grado de vinculación emocional de los clientes hacia la empresa.</p> <p>Referencia: Capitulo IV-Marco Practico, 7. Pre-Procesamiento y 8. Análisis de Sentimiento, Páginas 46-49.</p>

Objetivo	Tareas realizadas
<p>Objetivo específico 4</p> <p>Diseñar tableros de visualización de los datos e información para facilitar la comprensión e interpretación de los datos de las redes sociales.</p>	<p>Una vez consolidado el almacén de datos con la información de las redes sociales Facebook y Twitter, se ha procedido al diseño de tableros de visualización de los datos:</p> <p>Fase 3: Desarrollo</p> <ul style="list-style-type: none"> - Diseño de informes y cuadros de mando (dashboards) <ul style="list-style-type: none"> o Cuadros de mando para la red social Facebook. o Cuadros de mando para Twitter <p>Referencia: Capitulo IV-Marco Practico, 4.1.3 Desarrollo, e). Diseño de informes y cuadros de mando (dashboards), Páginas 68-75.</p>

Tabla 18 - Detalle de cumplimiento de objetivos
Fuente: *Elaboración propia*

4.2.2 Cuadro con el cumplimiento de la operacionalización de variables

La Tabla 19 detalla el cumplimiento de la operacionalización de variables definidas en el proceso de desarrollo del Modelo de inteligencia de negocios.

VARIABLE	DEFINICIÓN	DIMENSIONES	INDICADORES	ITEMS
<p>VARIABLE INDEPENDIENTE</p> <p>Modelo de Inteligencia de Negocios en redes sociales</p>	<p>La Inteligencia de Negocios, es entendido como el conjunto de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización [Conesa & Curto 2010], bajo este mismo concepto un modelo de Inteligencia de Negocios en Redes Sociales constituye la incorporación, procesamiento y exploración de los datos provenientes de las redes sociales a los procesos BI.</p>	<ul style="list-style-type: none"> - Procesos de extracción, transformación y carga de datos (ETL). - Almacén de datos (Datawarehouse) - API de acceso a datos de Twitter y Facebook. - Analítica de datos - Minería de Textos u opinión <p>Referencia: Capitulo IV-Marco Practico, Paginas 37-75.</p>	<ul style="list-style-type: none"> - Total seguidores (Twitter) - Nuevos seguidores de perfil (diarios, semanales, mensuales). - Menciones - Respuestas - Retuits - Total Fans (Facebook) - Total de “Me gusta” en página. - Nuevos “Me gusta” en página (diarios, semanales, mensuales). - Fuente de nuevos “Me gusta”. - Numero de Me gusta en publicaciones. <p>Referencia: Capitulo IV-Marco Practico, 4.1.1</p>	<ul style="list-style-type: none"> - Visibilidad y exposición de la empresa - Comunidad (Twitter) - Tweets vinculados a un producto/servicio. - Datos de los seguidores. - Comunidad (Facebook) - Publicaciones (Posts). <p>Referencia: Capitulo IV-Marco Practico, Cuadros de Mando. Páginas 68-75.</p>

VARIABLE	DEFINICIÓN	DIMENSIONES	INDICADORES	ITEMS
			Análisis de Requerimiento. Páginas 37-38.	
VARIABLE DEPENDIENTE Toma de decisiones empresariales	Toma de decisiones empresariales de forma óptima y oportuna.	<ul style="list-style-type: none"> - Análisis de opinión /sentimiento - Herramientas de Visualización de datos. - Cuadros de mando (Dashboards) - Reportes <p>Referencia: Capitulo IV-Marco Practico, Páginas 68-75.</p>	<ul style="list-style-type: none"> - Número de “Me gusta” hacia un producto y/o Servicio. - Número de compartidos. - Palabras más usadas para referirse a la empresa en los tweets y/o opiniones. - Palabras positivas. - Palabras negativas. <p>Referencia: Capitulo IV-Marco Practico, Páginas 69,70, 71 y 74.</p>	<ul style="list-style-type: none"> - Optimización del marketing. - Capturar ideas y clientes insatisfechos. - Cuáles son los más y/o menos favoritos. - Reputación de la institución (Nube de palabras) <p>Referencia: Capitulo IV-Marco Practico, Páginas 68-75.</p>

*Tabla 19 - Cumplimiento de operacionalización de variables
Fuente: Elaboración propia*

CAPÍTULO V: CONCLUSIONES Y RECOMENDACIONES DE LA INVESTIGACIÓN

5.1 Conclusiones y Recomendaciones

5.1.1 Conclusiones

Por lo expuesto y presentado en el presente documento se concluye que:

- Se ha cumplido con los objetivos planteados para el desarrollo del Modelo de Inteligencia de Negocios para Analizar el Posicionamiento de Productos y Servicios en Redes Sociales, de acuerdo al detalle de cumplimiento de objetivos especificado en el punto 4.2.1 Estado de los objetivos y resultados alcanzados.
- Con el desarrollo del presente trabajo de investigación hemos tenido la oportunidad de aplicar técnicas computacionales para la extracción, clasificación, comprensión, evaluación de opiniones, entre otros. No obstante, se concluye que uno de los puntos que ha generado más trabajo e investigación es extracción de datos de redes sociales, limpieza y pre procesamiento de datos.
- La visualización de datos es muy importante para la explotación de los datos y es una forma eficiente de encontrar patrones frecuentes o relaciones en un conjunto de datos. Es así que se ha procurado diseñar las representaciones más adecuadas para dar cumplimiento a los objetivos planteados, sin embargo, estos no son definitivos y se pueden explorar más técnicas de visualización para facilitar al usuario una mayor comprensión.

5.1.2 Recomendaciones

- El desarrollo de un modelo Inteligencia de Negocios implica la aplicación de diferentes herramientas, técnicas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización. Siguiendo el enfoque tradicional se compone de básicamente por los orígenes de datos, repositorio de datos y herramientas y técnicas de análisis y explotación, cada una de estas fases podría representar un objeto de estudio e investigación, puesto que cada fase implica un proceso diferente e independiente.

- Considerando a los sistemas de inteligencia de negocio son importantes para mejorar la toma de decisiones, se recomienda promover la investigación en soluciones de BI orientados tecnologías como: computación en la nube (cloud computing), big data y otras tendencias.
- A partir de los datos generados en la presente investigación puede servir de base para la aplicación de algoritmos de aprendizaje automático (machine learning) en la clasificación de mensajes y comentarios.

5.2 Anexos

Anexo 1 - Ejemplos de KPI para Redes Sociales

En Facebook	<ul style="list-style-type: none"> - Total de “Me gusta” en página - Nuevos “Me gusta” en página (diarios, semanales, mensuales). - Segmentación de los “Me gusta” - Fuente de nuevos “Me gusta” - Páginas sugeridas - En página - A través de publicaciones - Otras fuentes - Personas hablando del tema - Mensajes en el muro. - Engagement - “Me gusta” en publicaciones - Comentarios en publicaciones - Compartir en publicaciones - Total de visitas en Fanpage - Procedencia de visitas - Alcance de publicaciones - Clics en publicaciones - Clics en enlaces
En Twitter	<ul style="list-style-type: none"> - Total seguidores - Nuevos seguidores de perfil (diarios, semanales, mensuales). - Engagement. - Menciones - Respuestas - Retuits - Favoritos - Impresiones - Clics en enlaces - Listas en las que una cuenta aparece.

Anexo – 2. Herramientas de medición de datos de redes sociales

Herramientas de análisis	<ul style="list-style-type: none">- Google Analytics- Adobe Marketing Cloud
Herramientas de monitorización	<ul style="list-style-type: none">- Radian 6- Brandwatch- Talkwalker- Social Mention- Howsociable- Addictomatic
Herramientas de reporting	<ul style="list-style-type: none">- Facebook Insights- Twitter Analytics- Youtube Analytics- Google Plus- LinkedIn- Pinterest- Crowdbooste- TrueSocial - Metrics- Cyfe
Herramientas de benchmarking	<ul style="list-style-type: none">- SocialBackers- DKS Social Smart- Twittercounter- chengedetection.com- Compete Monkey- Wildfire Social Media Monitor
Herramientas de management	<ul style="list-style-type: none">- Tweetdeck- Hootsuite- Sprout Social- Buffer

5.3 Bibliografía

[Arias & Caminals 2013] Arias L. Emilio & Caminals S. Alex, “*Nuevas tendencias y Otros Procesos de BI*”, Edición FUOC 2013 pp. 27-32, Editorial Oberta UOC Publishing, SL, Barcelona - España.

[Conesa & Curto 2010] Conesa C. Jordy, Curto D. Josep. “*Introducción al Business Intelligence*”, 2010, Editorial UOC. España.

[Aguilar 2003] Aguilar Q. Ramiro, “*Minería de Datos, Fundamentos, Técnicas y Aplicaciones*”, 2003, Editorial Reprografía Signo, España.

[Bartroli, Braulio, Curto & Puigvert 2015] Bartroli M. Alex, Braulio G. Nuria, Curto D. Josep, Puigvert P. Francesc, “*Nuevas Tendencias Tecnológicas en BI*”, Primera Edición 2015 pp. 7-14, Editorial Oberta UOC Publishing, SL, Barcelona - España.

[Bartroli, Braulio, Curto & Puigvert 2015] Bartroli M. Alex, Braulio G. Nuria, Curto D. Josep, Puigvert P. Francesc, “*Nuevas Tendencias Tecnológicas en BI*”, Primera Edición 2015 pp. 7-14, Editorial Oberta UOC Publishing, SL, Barcelona - España.

[Arias & Caminals 2013] Emilio Arias Leal & Alex Caminals Sánchez de la Campa, “*Nuevas tendencias y Otros Procesos de BI*”, Edición FUOC 2013 pp. 27-32, Editorial Oberta UOC Publishing, SL, Barcelona - España.

[Gironés 2013] Gironés R. Jordi, “*Business Analytics*”. Edición FUOC, 2013 pp. 8-17, Editorial Eureka Media, SL, Barcelona - España.

[Gironés Roig 2013] Gironés Roig Jordi “*Data Mining*”. Edición FUOC, 2013 pp. 5, Editorial Eureka Media, SL, Barcelona - España

[Joeri Heijnen-2012] Tesis de Maestría titulada: *Social Business Intelligence - How and where firms can use social media data for performance measurement, an exploratory study*, Universidad Tecnológica de Delft, Nederland, 2012.

[Elena Asensio Blasco - 2015] Tesis de Maestría titulada: *Aplicación de técnicas de minería de datos en redes sociales/web. Universidad Politécnica de Valencia, España, 2015.*

[Wikipedia Red Social 2019] *Red Social*, Información en línea. Disponible en: https://es.wikipedia.org/wiki/Red_social. Consulta: 25/04/2019.

[Wikipedia Servicio Red Social 2019] *Servicio de Red Social* https://es.wikipedia.org/wiki/Servicio_de_red_social. Consulta: 01/05/2019.

[Wikipedia Facebook 2019] Facebook, información en línea. Disponible en: <https://es.wikipedia.org/wiki/Facebook>. Consulta 01/05/2019.

[Wikipedia Twitter 2019] Twitter, Información en línea. Disponible en: <https://es.wikipedia.org/wiki/Twitter>. Consulta 01/05/2019.

[Curto 2009] Curto Díaz Josep, “*Introducción a Business Intelligence*”. Presentación en línea publicada el 18/12/2009. Disponible en: <http://es.slideshare.net/josep.curto/introduccion-a-business-intelligence>. Consulta: 25/04/2019.

[Wikipedia Minería de Datos 2019] Minería de Datos, información en línea. Disponible en: https://es.wikipedia.org/wiki/Minería_de_datos. Consulta 01/03/2019.

[Latamclick 2018] Estadísticas de Facebook 2018 en América Latina. Disponible en: <https://www.latamclick.com/estadisticas-de-facebook-america-latina-2018/>. Consulta: 01/05/2019.

[ARTICULO XII], Informe sobre Control Estatal de las Redes Sociales, <http://www.alianzaregional.net/wp-content/uploads/Informe-Articulo-XIII-2016-GF-SR-DM.pdf>. Consultado en 15/09/2020.