

Análisis de Información de Redes Sociales (Twitter)

Angélica Urrutia¹ y Carolina Nicolás²

¹Universidad Católica del Maule Talca Chile

²Universidad Santo Tomás Santiago Chile

RESUMEN

El vertiginoso desarrollo de la tecnología y con ello la globalización del conocimiento, ha generado un alto interés de las redes sociales en las organizaciones, y su presencia se ha multiplicado exponencialmente en los últimos años. Es por ello, que aquí se propone analizar los datos extraídos de redes sociales, específicamente Twitter, con la finalidad de obtener diferentes elementos que permitan la gestión y el análisis relacionado con las opiniones que los usuarios proporcionan sobre las distintas empresas que utilizan a Twitter como una herramienta social en sus páginas web. Esta información es de gran utilidad para la gestión de clientes y su preferencia a marcas u organizaciones.

CONTEXTO

Llevar un registro de las opiniones y comentarios que se están compartiendo en la red y así dar respuestas directas a dudas o críticas de los clientes, o bien analizar los datos, o promocionar productos, realizar estudios de mercado, entre otras opciones es fundamental para las organizaciones [1,5, 10].

Propuesta perfil del usuario.

El análisis de los datos extraídos de Twitter, de esta investigación, es focalizado en perfiles de organizaciones: bancarias, telefónicas, retail y supermercados, seleccionando las cuentas vigentes. El interés de la información extraída de cada página se centra en la *visualización del perfil del tipo de tuitero*, considerando las siguientes características:

- **Liderazgo**, entendiéndose con ello que se desea visualizar qué tweet es el que se está hablando con más recurrencia y a quién pertenece. Para ello es necesario contar con los mensajes emitidos y con información como **la cantidad de “retweet”** y **“Me gusta”** que poseen los tweets que hacen referencia a los diferentes perfiles.
- **Experiencia**, este enfoque busca visualizar que tan participativo o experto es el usuario que realiza el tweet, ya sea porque posee gran cantidad de mensajes emitidos, alta cantidad de personas a las que les interesa su perfil, por la cantidad de perfiles de interés o la cantidad de **“Me gusta”** que realiza a diferentes tweets. Para esto, la información que se requiere es la relacionada con el perfil del **usuario**, donde los datos relevantes para este análisis corresponden a la **fecha de creación de dicha cuenta, la cantidad de tweets, cantidad de followers, cantidad de following** y **cantidad de likes** que posee el usuario.
- **Origen**, que utiliza la información relacionada con la **localidad** del usuario, que es ingresada por él mismo al momento de crear su cuenta. Esto restringe a obtener un análisis efectivo, ya que la localidad se presenta en diversos formatos o incluso de manera incorrecta, así como también debido a que no es información obligatoria, nos encontramos con usuarios sin definir su localidad.

A modo de resumen, se presenta a continuación la Figura 1, que muestra un diagrama con los requerimientos mencionados y las características requeridas en el análisis del perfil del usuario de una red social, caso Twitter.

Para el análisis de los datos, se considera la recopilación de diferentes tweets de las páginas web de las empresas seleccionadas [2, 7], se extraen datos asociados a las características para cada perfil, y posterior implementación utilizando la herramienta, como por ejemplo Qlik Sense Desktop, con la finalidad de generar visualizaciones personalizadas e interactivas, dashboard, de datos que busquen responder las siguientes interrogantes de investigación:

- ¿Cómo el tuitero se comporta en las diferentes marcas?
- ¿Cómo afectan las opiniones en la red, al perfil de Twitter de la empresa?

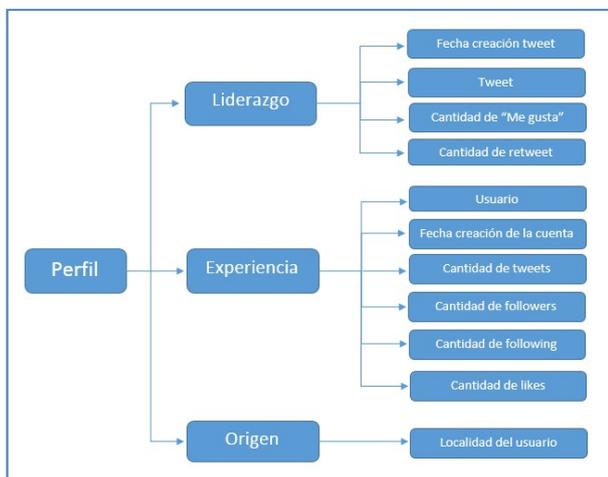


Figura 1: Características propuestas para perfil del usuario Twitter.

1. INTRODUCCIÓN

Para la implementación y obtención de análisis de resultados del perfil del usuario, es de vital importancia la existencia de una arquitectura de software definida adecuadamente, que permita extraer y analizar datos desde Twitter [3, 4, 9, 12].

Propuesta de arquitectura para analizar perfil twitter.

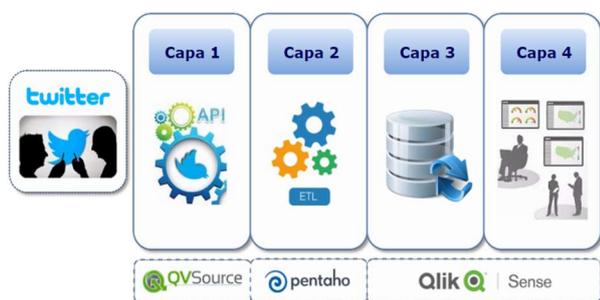


Figura 2: Propuesta de arquitectura de software para analizar perfil Twitter.

La Figura 2 muestra una propuesta de una arquitectura de software/herramientas compuestas por cuatro capas de procesos descritas a continuación, cada una de ellas requiere de herramientas específicas asociada

a diferentes tareas y procesos, como por ejemplo: QVSource, Pentaho y Qlik Sense Desktop.

Capa 1: Extracción de la fuente de datos. Es la primera capa la que realiza la extracción de datos desde Twitter haciendo referencia a los tres perfiles propuestos. Para el proceso de extracción se utiliza la herramienta QVSource, que permite conectar Qlik Sense Desktop a la gran cantidad de APIs sociales y empresariales que se encuentran disponibles en la web. En este caso, se utiliza QVSource como conector con la API Search de Twitter de forma interactiva, permitiendo así obtener los datos de todos los tweets que hacen referencia a cada usuario e información específica del usuario que realiza dichos tweets.

Capa 2: Procesos ETL. En esta segunda capa se utiliza la herramienta Spoon de Pentaho Data Integration, para llevar a cabo los procesos de transformación y carga de los datos. Los datos obtenidos, son transformados con el propósito que sean lo más claro y limpios posible, para luego generar un solo archivo con los datos de los doce perfiles seleccionados para su posterior análisis.

Capa 3: Selección base de datos. Esta capa adquiere importancia, ya que, en ella es donde se comienza a utilizar el software interactivo Qlik Sense Desktop, para el análisis del perfil de usuario Twitter. Como primera tarea, se carga el archivo generado en la Capa 2, luego seleccionar los atributos o datos de interés, para llevar a cabo el modelo de análisis de la información, ya propuesto en la Figura 1. De esta forma se obtiene la base de datos con los elementos de análisis que se requieren.

Capa 4: Visualización de los resultados. En esta última capa es donde se visualizan los resultados obtenidos desde las capas anteriores de la arquitectura propuesta. A partir del análisis de la base de datos generada y las diferentes opciones de visualización que entrega Qlik Sense Desktop, se construyen dashboard dinámicos y de fácil comprensión, con la finalidad de dar respuestas a los indicadores de gestión propuestos de manera

que sean útiles para la toma de decisiones con respecto a lo que opinan los tuiteros [6, 8, 11].

Producto de esta investigación.

Se tiene una arquitectura que permite realizar la implementación para el análisis de datos extraídos desde herramientas sociales (Twitter), relacionado con el valor de la experiencia de clientes, a través del modelo de “perfil de usuario”, para lo cual se utiliza como herramienta de visualización de datos, el programa QlikSense Desktop. Este tipo de solución presenta una gran ventaja que es su bajo costo y, además, posee la particularidad que puede ser utilizada, por cualquier persona que tenga la necesidad de resolver este modelo para el desarrollo y optimización de su organización, sin la necesidad de contar necesariamente con expertos en el área de Inteligencia de Negocios.

La solución propuesta, cumple con las etapas fundamentales de un modelo de Inteligencia de Negocios, puesto que se proporciona un modelo de análisis de información, diferentes visualizaciones, que permiten entregar respuestas a las inquietudes planteadas por las empresas para tomar las mejores decisiones basadas en información relevante.

Obtener información que proporcionan los usuarios a través de las redes sociales, constituye un elemento de alto valor y un punto fundamental para las organizaciones, ya que es a través de ellas donde las personas exponen reclamos, opiniones, experiencias o necesidades que son relevantes para ellos y que constituye una información valiosa para todo tipo de entidades que deseen permanecer conectadas con el mercado.

Para llevar a cabo este análisis de información, la red social tomada como ejemplo es Twitter, ya que, es una de las redes más utilizadas, con una gran cantidad de usuarios activos. Por esta razón, la elección de esta red para desarrollar el modelo de perfil de usuario se consideró como la más adecuada e ideal, principalmente por su característica particular de poseer mensajes de texto limitados a 140 caracteres, y gracias a que la información entregada por Twitter es de acceso público, al igual que las aplicaciones

que permiten conectarse a la API de esta red social para así obtener la información deseada. De esta manera, la posibilidad de acceder a datos generados por el mundo exterior a través de las redes sociales, está al alcance de cualquier organización, procurando así una gran y valiosa oportunidad para ellas de acercarse en forma adecuada y con información relevante a sus clientes, ya que esta posibilidad les ha permitido conocerlos con mayor profundidad.

Esta alternativa, es hoy día crucial en el mundo globalizado y con organizaciones y clientes cada vez más exigentes e informados. Por ello, las organizaciones se ven enfrentadas de manera permanente a desafíos, en que la toma de decisiones que sean correctas puede hacer la diferencia entre lograr el éxito o simplemente el fracaso. En este sentido, el acceso a información útil, relevante y confiable permite a la empresa tomar decisiones estratégicas y orientar sus operaciones de negocio enfocándose en la gestión comercial de sus clientes que impactará positivamente en la imagen corporativa de la empresa y con ello, ser competitiva y consolidar su posición en el mercado.

2. LINEAS DE INVESTIGACIÓN Y DESARROLLO

En el transcurso de la investigación realizada se han identificado algunos temas interesantes que pueden surgir a partir de esta. A continuación se proponen algunos trabajos futuros que se pueden realizar.

Algoritmo para realizar análisis de sentimiento de los tweets: Esta investigación entrega la base para obtener diferentes análisis del valor de la experiencia de clientes de Twitter, el cual puede ser complementado relacionado otros factores asociados a los mensajes para así obtener estudios más específicos.

Gestión de la Información: Con los datos almacenados por cada tweet se puede focalizar preguntas de investigación más específicas, asociadas a la gestión de satisfacción de marca de cada cliente que

accede a la página web de las organizaciones y entrega su opinión en ella. Una tarea pendiente es ampliar las herramientas sociales como facebook entre otras.

3. RESULTADOS OBTENIDOS O ESPERADOS

El análisis de los datos y sus representaciones gráficas son relevantes en esta investigación, es así como la Figura 3 muestra el dashboard, que da respuesta a la experiencia y localidad indicados en la propuesta de perfil del usuario.



Figura 3: Dashboard en relación al enfoque de la experiencia.

El dashboard permite complementar información entre las visualizaciones, siendo un tema relevante la fecha de creación de la cuenta del usuario, de esta forma se analiza la cantidad de tweets, likes, followers y following, que el usuario posee durante el tiempo que lleva interactuando en Twitter. Por ejemplo, no es lo mismo un usuario con miles de seguidores que creó su cuenta hace 10 años atrás, que un usuario con la misma cantidad de seguidores incorporado a la red social hace un par de años, pues tiene mayor participación y popularidad este último que posee una gran cantidad de usuarios a quienes le interesa su perfil en un periodo más corto. Para un análisis más completo se utiliza la información que indica la cantidad de tweets que el usuario ha realizado, visualizando de esta manera los usuarios que tienen mayor participación en la red social en un corto plazo.

A pesar de que la mayoría de los usuarios no ingresa su localidad o lo hace de manera

incorrecta, se decide de igual manera, realizar un análisis donde se visualiza la cantidad de tweets realizados por la localidad del usuario, para tener una mirada de esta información y generar la posibilidad de analizar si será de utilidad según las necesidades. Para este dashboard, se ha utilizado un panel de filtrado, que permite visualizar la información, según el año en que el usuario creó su cuenta de Twitter, y también más específicamente, por día y mes de la creación de dicha cuenta. Finalmente, la información y análisis de datos tienen un fin de utilidades y formar de visualizar que apoyarán la gestión de las empresas en base al perfil de usuario propuesto en la Figura 1..

4. FORMACION DE RECURSO HUMANO

En esta investigación trabajan académicas de la Universidad Santo Tomás en Chile con el proyecto “Twitted, its effect on the experience value of the Brand”, y académicas de la Universidad Católica del Maule con el grupo de investigación en bases de datos TRICAHUE.

En Ambas universidades participan alumnos de pregrado y postgrado de las carreras de Ingeniería Comercial e Ingeniería Civil Informática, desarrollando tesis de título y grado de licenciado y magíster.

5. BIBLIOGRAFIA

- [1] GHIASSI, M.; SKINNER, J.; ZIMBRA, D. (2013). Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network. *Expert Systems with applications*, vol. 40, no 16, p. 6266-6282.
- [2] ESULI, F. SEBASTIANI (2006): Determining term subjectivity and term orientation for opinion mining. In Proceedings of EACL-06, 11th Conference of the European Chapter of the Association for Computational Linguistics, pp. 193-200.
- [3] JANSEN, B.J.; ZHANG, M.; SOBEL, K. (2009): Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*, Vol. 60, No. 11, pp. 2169-2188.
- [4] KAZUSHI IKEDA, GEN HATTORI, CHIHIRO ONO, HIDEKI ASOH, TERUO HIGASHINO

- (2013). Twitter user profiling based on text and community mining for market Analysis Japon. Knowledge-Based Systems Volume 51, Pages 35-47.
- [5] LIN J., RYABOY D. (2012): Scaling big data mining infrastructure: the twitter experience. Newsletter ACM SIGKDD Explorations Newsletter Volume 14 Issue 2. pp. 6-19 New York USA.
- [6] LEWIS D.D. (1998). Naive (Bayes) at forty: The Independence assumption in information retrieval. In Proceedings of ECML-98, 10th European Conference on Machine Learning, Chemnitz, Germany, pp. 4-15.
- [7] MARTÍNEZ E., M. MARTÍN T., M. PEREA J. M., L. UREÑA A. (2011): Técnicas de clasificación de opiniones aplicadas a un corpus en español. Procesamiento del Lenguaje Natural, Revista nº 47 septiembre de 2011, pp. 163-170. Universidad de Jaén.
- [8] MOSTAFA, MOHAMED M. (2013). More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications*, vol. 40, no 10, p. 4241-4251.
- [9] NIGAMY K., LAERTY J., MCCALLUMZ A. (1999). Using Maximum Entropy for Text Classification. In IJCAI-99 Workshop on Machine Learning for Information Filtering.
- [10] STRATEBI (2013). Nuevas tendencias en Bussiness Intelligence del Big Data al Social Intelligence.
http://www.stratebi.es/todobi/May13/Nuevas_Tendencias_BI.pdf Madrid, España.
- [11] SAXENA, A., & GADHIYA, S. (2014). A Survey on frequent pattern mining methods-Apriori, Eclat, FP growth. In *International Journal of Engineering Development and Research* (Vol. 2, No. 1) IJEDR.
- [12] SEBASTIANI F. (2002). Machine Learning in Automated Text Categorization. *ACM Computing Surveys*, 34-1, pp. 1-47.