Twitter como herramienta para medir la percepción de los consumidores finales en una entidad de bienes de consumo.

Jorge Eduardo Silva Páez

Alexander Peña Bautista

Jaime Enrique Plazas Martínez

Eder Orlando Moreno Villanueva

Especialización Big data, Fundación Universitaria Compensar

Trabajo de grado presentado como requisito parcial para optar al título de:

Especialización Big data.

Omar Alexander Leon Garcia

19 de mayo de 2023

Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional.

Twitter como herramienta para medir la percepción de los consumidores finales en una entidad de bienes de consumo.

Jorge Eduardo Silva Páez
Alexander Peña Bautista
Jaime Enrique Plazas Martínez
Eder Orlando Moreno Villanueva

Trabajo de grado presentado como requisito parcial para optar al título de:

Especialización Big data.

Director (a):
Jairo Cesar Gomez Acero
Codirector (a):

Línea de Investigación:

Análisis de sentimientos y minería de datos en redes sociales.

Fundación Universitaria Unipanamericana – Compensar
Facultad de Ingeniería, Ingeniería de (Especialización en Big Data)
Bogotá, Colombia
2023

Resumen

El presente estudio, se enfoca en mejorar la capacidad de análisis de la percepción de los consumidores finales de una entidad de bienes de consumo mediante el uso de la red social Twitter. El objetivo del proyecto es desarrollar técnicas avanzadas de análisis de sentimiento y minería de datos para extraer información relevante de los tweets de los usuarios, comprender cómo se sienten los consumidores respecto a la entidad en cuestión y utilizar esta información para generar herramientas que puedan ayudar a mejorar la toma de decisiones.

El proyecto involucra el uso de un método que realiza el análisis de sentimiento para identificar las opiniones y las emociones que los consumidores expresan en los tuits relacionados con la entidad de bienes de consumo. Se utilizan técnicas de machine learning tales como aprendizaje no supervisado, el cual por medio de un método se envían datos sin etiquetar y se encarga entender los datos y determinar el sentimiento.

Palabras clave: Twitter, minería de datos, aprendizaje automático, Tuit.

Abstract

The main focus of this study is improve the ability to analyze the perception of end consumers that belong to a goods entity through the use of the social network Twitter. The objective of this project is to develop advanced sentiment analysis and data mining techniques to extract relevant information from users' tweets, understand how consumers feel about the entity in question, and use this information to generate tools that can help improve decision making.

The project involves the use of a method that performs sentiment analysis to identify the opinions and emotions that consumers express in tweets related to the consumer goods entity. Machine learning techniques such as unsupervised learning are used, which through a method that sends unlabeled data and is responsible for understanding the data and determining the sentiment.

Keywords: Twitter, data mining, machine learning, Tweet**.**

Contenido

Pág.

Resumen 3
Contenido 5
Lista de Ilustraciones. 6
Introducción 7
Antecedentes y Justificación 8
Formulación del Problema 10
Objetivos 12
Objetivo General 12
Objetivos Específicos 12
Alcances y Limitaciones 13
Alcances 13
Limitaciones 13
1. Marco teórico 14
1.1. Trabajos relacionados 25
2. Diseño metodológico 26
2.1. Fase 1: Análisis 26
2.2. Fase 2: Diseño e integración: 26
2.2.1. Obtención de credenciales de acceso: 27
2.2.2. Autenticación en la API de Twitter: 27
2.2.3. Recopilación de datos: 27
2.2.4. Análisis de los datos: 27
2.2.5. Representación visual de los datos: 27
2.2.6. Interpretación de resultados: 27
2.3. Fase 3: Implementación de las herramientas visuales: 28
3. Resultados 28
3.1. Diseño de la solución. 28
3.2. Combinación de herramientas visuales. 34
3.3. Discusión De Resultados 36
4. Conclusiones y recomendaciones. 37
4.1 Conclusiones 37

4.2. Recomendaciones 39
Referencias 40

Lista de Ilustraciones.

Pág.

Ilustración 1. Logo Twitter. 17
Ilustración 2. Tweet 17
Ilustración 3. Logo Twitter Developers 18
Ilustración 4. Logo Twitter API 18
Ilustración 5. Logo Python. 19
Ilustración 6. Logo librería Tweepy 20
Ilustración 7. Logo Procesamiento de Lenguaje Natural. 20
Ilustración 8. Logo librería Textblob. 21
Ilustración 9. Logo visual Studio Code 22
Ilustración 10. Logo librería Seaborn. 22
Ilustración 11. Logo librería WordCloud. 23
Ilustración 12. Logo machine learning. 23
Ilustración 13. Logo aprendizaje no supervisado. 24
Ilustración 14. Diseño de la solución 28
Ilustración 15. Grafica resultado 1. 29
Ilustración 16. Grafica resultado 2. 30
Ilustración 17. Grafica resultado 3. 31
Ilustración 18. Grafica resultado 4. 32
Ilustración 19. Grafica resultado 5. 32
Ilustración 20. Grafica resultado 6. 33
Ilustración 21. Tuits Negativos 34
Ilustración 22. Combinación análisis negativos. 34
Ilustración 23. Tuits positivos 35
Ilustración 24. Combinación análisis positivos 35

Introducción

En la actualidad, las redes sociales han cambiado la forma en que las empresas interactúan con sus consumidores y cómo éstos perciben los productos o servicios que se ofrecen. Twitter es una de las principales redes sociales, ya que millones de usuarios realizan tuits a diario. Por lo tanto, analizar la percepción de los consumidores finales mediante el uso de Twitter se ha convertido en una herramienta esencial de las empresas para mejorar la relación con sus clientes y tomar decisiones empresariales informadas (Carmona, 2016).

Sin embargo, analizar grandes cantidades de tweets puede resultar complejo y dispendioso para las empresas. Es por eso, por lo que se hace necesario desarrollar técnicas avanzadas de análisis de sentimiento y minería de datos para extraer información relevante de los tweets de los usuarios y comprender cómo se sienten los consumidores acerca de la entidad en cuestión (Kuchinke, 2010).

Este proyecto de investigación en Big Data tiene como objetivo principal mejorar la capacidad de análisis de la percepción de los consumidores finales de una entidad de bienes de consumo mediante el uso de la red social Twitter. Para ello, se llevará a cabo una exploración de diversas metodologías de análisis de sentimiento para identificar las opiniones y emociones que los consumidores expresan en los tweets relacionados con la entidad de bienes de consumo (Kuchinke, 2010). Además, se generarán herramientas para mejorar el análisis de los consumidores una vez se hayan identificado estas opiniones y emociones.

De acuerdo con lo anterior, se entregarán herramientas que proporcionarán una mejora, en el momento de analizar la información por parte de las entidades, esto con la finalidad de tomar mejores decisiones.

Antecedentes y Justificación

El estudio propuesto será útil para empresas que buscan mejorar la calidad de sus productos y servicios a través de la comprensión de las opiniones de los consumidores en la red social Twitter. Este estudio podría contribuir al desarrollo de herramientas y estrategias para el análisis de la percepción de los consumidores en Twitter, lo que podrá mejorar la reputación y la satisfacción del cliente de las empresas. También podrá ayudar a las empresas a identificar áreas de mejora en sus productos y servicios(Wu et al., 2017).

El análisis de percepción que se plantea realizar contribuirá a entender la imagen de la empresa, a visualizar como están siendo percibidos en las redes sociales, de esta manera se pueden tomar decisiones estando informados lo cual contribuye a identificar problemas y así poder formular estrategias con el fin de obtener una mejora y un mejor posicionamiento frente a la competencia. Siendo fundamental para cualquier entidad de bienes de consumo, esto va a permitir conocer cómo son percibidos sus productos o servicios por parte de los consumidores y qué aspectos se pueden mejorar para satisfacer mejor sus necesidades y expectativas. El uso de la plataforma Twitter se ha convertido en una herramienta muy útil para realizar el análisis, ya que va a permitir información en tiempo real sobre las opiniones, comentarios y sugerencias de los usuarios acerca de los productos o servicios de las empresas (M. R. González-Rodríguez, 2015).

En la actualidad, las redes sociales se han convertido en una herramienta fundamental para las empresas que desean mantenerse competitivas en el mercado, ya que les permiten conocer la opinión de los consumidores acerca de sus productos y servicios. Específicamente, Twitter se ha posicionado como una plataforma que facilita la interacción entre las empresas y los consumidores, ofreciendo información en tiempo real sobre sus percepciones y necesidades. Por esta razón, resulta relevante para las empresas conocer cómo pueden utilizar esta herramienta

para mejorar la calidad de sus productos y servicios. Además, el análisis de las redes sociales contribuirá a la identificación de los aspectos que deben mejorarse en los productos y servicios ofrecidos, lo que redundará en la satisfacción del cliente y en la mejora de la reputación empresarial. En este contexto, se hace necesario un estudio que permita comprender la relevancia de las redes sociales para las empresas, así como identificar las estrategias que pueden ser implementadas para mejorar la calidad de los productos y servicios ofrecidos(M. R. González-Rodríguez, 2015).

Formulación del Problema

Actualmente las redes sociales se han convertido en herramientas muy valiosas, para las empresas y personas, que buscan conectarse con sus clientes y conocer sus opiniones y diferentes percepciones sobre sus productos, servicios etc… La plataforma Twitter permite a las personas y empresas interactuar directamente con los consumidores y poder obtener información útil sobre sus gustos y preferencias (Brunetta, 2013).

Las empresas han intentado resolver diferentes problemas, como la implementación de encuestas de satisfacción al cliente, la realización de pruebas de productos para evaluar la calidad de sus productos. Las empresas enfocan su objetivo en la plataforma Twitter, ya que son una fuente valiosa de información, donde los consumidores pueden compartir sus experiencias, quejas, solicitudes y opiniones de manera abierta. Sin embargo, no todas las empresas saben cómo analizar los diferentes tipos de información que recopila la plataforma Twitter y cómo utilizarla para mejorar sus productos y servicios (Brunetta, 2013).

Twitter es una de las redes sociales más utilizadas por empresas y consumidores para interactuar y compartir información. Las empresas han encontrado en esta plataforma una valiosa herramienta para conocer la opinión de sus clientes y adaptar sus productos y servicios a las necesidades y preferencias de los consumidores.

En la actualidad, las opiniones de los consumidores son cada vez más importantes para las empresas, ya que estas opiniones pueden influir en la toma de decisiones y en la reputación de las mismas en el mercado. A medida que las redes sociales se han vuelto más populares, se ha visto un aumento en el uso de estas plataformas para expresar opiniones sobre productos y servicios. Esto ha llevado a las empresas a tratar de comprender mejor cómo los consumidores

perciben sus productos y servicios en la web social (Tausczik & Pennebaker, 2010). En este sentido, surge la pregunta:

¿Cómo mejorar la capacidad de análisis de la percepción de los consumidores finales de una entidad de bienes de consumo mediante el uso de la red social Twitter?

Objetivos

Objetivo General

Mejorar el análisis de la percepción de los consumidores finales de una entidad de bienes de consumo utilizando la información extraída de la red social Twitter.

Objetivos Específicos

- Diseñar e implementar un modelo que permita realizar la clasificación de los comentarios realizados en Twitter relacionados a una entidad en específico.
- Evaluar la percepción de las personas en la red social Twitter, para determinar si se presenta una calificación por debajo de los limites esperados.
- Generar herramientas que permitan realizar el análisis de los sentimientos de los consumidores finales en relación con los productos y servicios de una empresa.

Alcances y Limitaciones

Alcances

Generar herramientas que permitan mejorar el análisis de sentimientos en Twitter, que sea accesible a las entidades de bienes y servicio de acuerdo con el tipo de información que desean; analizar, identificar, evaluar y desarrollar.

Los entregables que se van a suministrar son:

Diseño de un modelo realizado bajo el código Python, que pueda ser implementado para cualquier entidad de bienes y consumo, para generar herramientas que permitan mejorar el análisis de los sentimientos de los consumidores finales de la empresa.

Limitaciones

Aunque los servicios de Twitter para extraer información funcionan de manera gratuita se hace necesario tener una cuenta autorizada para extraer información de manera masiva máximo 1500 tuits.

Si se requiere obtener una mayor cantidad de información se debe adquirir un nivel superior de cuenta que podría generar costos adicionales.

1. Marco teórico

Las redes sociales se han convertido en el medio de comunicación de preferencia de las personas, por ende, las empresas se han visto en la obligación de cambiar sus estrategias de hacer negocios, así como sus ideas de hacer marketing; esto para poder tener una ventaja competitiva, de esta manera las empresas cada día están más interesadas en saber cómo están siendo percibidas por los clientes y usuarios. Cada vez recurren en interpretar más interacciones realizadas por ellos. La red social Twitter tiene una serie de ventajas, ya que permite obtener información en tiempo real de lo que está ocurriendo y obtener temas que son tendencia en el momento, logrando acceder la información mediante las herramientas proporcionadas por la aplicación, con la ventaja que es una información estructurada y con una gran cantidad de datos que pueden analizarse para que los usuarios puedan comprender y pueda ser utilizada por empresas, para mejorar sus estrategias de comunicación e interactuar con su audiencia de manera más efectiva. (Muñoz-Expósito, Oviedo-García, & Castellanos-Verdugo, 2017)

La percepción de los consumidores finales es crucial para cualquier entidad de bienes de consumo, ya que puede afectar directamente su éxito en el mercado. En los últimos años, las redes sociales han comenzado a desempeñar un papel importante en la recopilación de información sobre la percepción de los consumidores. En particular, Twitter se ha convertido en una plataforma popular para la recopilación de datos sobre la percepción de los consumidores(Guo & Chen, 2019). Según un estudio de(Hennig-Thurau et al., 2013), las redes sociales, incluyendo Twitter, son una fuente rica de información sobre la percepción de los consumidores. Los autores señalan que las redes sociales han permitido a las empresas recopilar datos en tiempo real y analizar la percepción del consumidor de manera más eficiente que las técnicas de investigación tradicionales. A través del análisis de los datos de Twitter, las empresas

pueden obtener información valiosa sobre las opiniones de los consumidores sobre sus productos y servicios(Hennig-Thurau et al., 2013).

Para mejorar la capacidad de análisis de la percepción de los consumidores finales a través de Twitter, es importante utilizar técnicas de análisis avanzadas. Según un estudio de (Fan, Gordon, & Pathak, 2019), el análisis de sentimientos es una técnica popular para analizar los datos de Twitter. Los autores señalan que el análisis de sentimientos puede ayudar a las empresas a comprender la actitud general de los consumidores hacia sus productos y servicios. Además, el uso de técnicas avanzadas de análisis de datos, como el aprendizaje automático, puede permitir a las empresas identificar patrones y tendencias en la percepción de los consumidores (Fan, Gordon, & Pathak, 2019).

Otra forma de mejorar la capacidad de análisis de la percepción de los consumidores a través de Twitter es mediante la segmentación de la audiencia. Según un estudio de (Guo & Chen, 2019), la segmentación de la audiencia puede ayudar a las empresas a comprender mejor las necesidades y deseos de diferentes grupos de consumidores. Al utilizar técnicas de segmentación de la audiencia, las empresas pueden personalizar su comunicación y marketing para satisfacer las necesidades específicas de cada grupo de consumidores, lo cual permitirá interactuar en campañas efectivas que puedan resultar de gran ayuda para las entidades de bienes y servicios(Guo & Chen, 2019).

Además, para mejorar la capacidad de análisis de la percepción de los consumidores a través de Twitter, es importante utilizar herramientas de monitoreo y análisis de redes sociales. Según un estudio de (Kietzmann, Hermkens, McCarthy, & Silvestre, 2011), las empresas pueden utilizar herramientas de monitoreo y análisis de redes sociales para recopilar y analizar datos sobre la percepción de los consumidores en tiempo real. Las herramientas de monitoreo y

análisis de redes sociales también pueden ayudar a las empresas a identificar y abordar problemas de manera más rápida y eficiente (Kietzmann, Hermkens, McCarthy, & Silvestre, 2011).

Para mejorar la capacidad de análisis de la percepción de los consumidores finales de una entidad de bienes de consumo mediante el uso de la red social Twitter implica utilizar técnicas avanzadas de análisis, segmentar la audiencia, y utilizar herramientas de monitoreo y análisis de redes sociales. Al realizar el proceso, las empresas pueden obtener información valiosa sobre las opiniones de los consumidores sobre sus productos y servicios, y utilizar esta información para mejorar su éxito en el mercado, poder atraer más servicios, entrar en la competencia del marketing, de acuerdo con el tipo de información que maneje la empresa y poder obtener más ideas de fortalecimiento para poder mejorar el estatus empresarial (Hennig-Thurau, y otros, 2013).

El análisis de la percepción es el mecanismo individual que realizan los seres humanos que consiste en recibir, interpretar y comprender las señales que provienen desde el exterior, codificándolas a partir de la actividad sensitiva. Se trata de una serie de datos que son captados por el cuerpo a modo de información bruta, que permitirá un significado, luego de un proceso cognitivo que también es parte de la propia percepción. Se denomina una técnica, que permitirá a las empresas, organizaciones a poder percibir, clasificar los tuits buenos y malos, recopilar las diferentes fuentes de información para poder identificar tendenticas, patrones de usuarios para entidades de bienes y servicios (Suarez, 2018).

Ilustración 1**.** Logo Twitter.



Fuente: (https://twitter.com/, s.f.)

Twitter es una red social de microbloging, fue diseñada como una plataforma social; para poder facilitar la circulación de mensajería en grupos, donde se puede compartir contenidos mediante publicaciones, pero con una limitación de 280 caracteres. Esta característica, más que una limitación es una ventaja, ya que permite consumir información de forma liviana y poder visualizar gran cantidad de publicaciones. La red social Twitter, permite que los usuarios decidan que escuchar, para poder hacer una delimitación en el círculo de quien interese. El contenido de la plataforma Twitter sigue transformando de una manera orgánica, significa que los usuarios deben experimentar y apropiarse de la ductilidad del sistema (Orihuela, 2011).

Ilustración 2**.** Tweet.

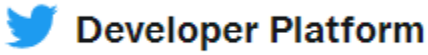


Fuente: (InfoMoney., 2020)

Tuit o tweet, es una publicación subida a Twitter con un máximo de 280 caracteres. En ellos se puede insertar Hashtags, nombrar usuarios, insertar otros tuits, insertar imágenes, incluir

links de artículos, publicaciones, insertar vídeos, audios, el cual tiene como objetivo la interacción de usuarios, para estrategias de marketing. Existen varios tipos de tuits como compartir links de páginas o perfil de los usuarios, poder establecer encuestas de satisfacción de productos de servicio, un seguimiento de un tema por medio de un hashtag seguido del tema, noticias relevantes, que es muy importante para el seguimiento del acontecimiento de interés (Borras, 2020).

Ilustración 3**.** Logo Twitter Developers.



Fuente: (TWITTER, s.f.)

Twitter Developers es una plataforma que proporciona herramientas y recursos para que los desarrolladores puedan crear aplicaciones y servicios basados en la API de Twitter. Adicional ofrece documentación, tutoriales, guías de inicio rápido, ejemplos de código y herramientas para ayudar a los desarrolladores a construir aplicaciones de Twitter, esto ayudara a realizar la extracción de la información de la plataforma de Twitter. Para la utilización de la API se debe crear una cuenta para desarrolladores dentro de la plataforma (TWITTER, s.f.).

Ilustración 4**.** Logo Twitter API.



Fuente: (TWITTER, s.f.)

Twitter proporciona a las empresas, los desarrolladores y los usuarios acceso programático a los datos mediante las API "Application Programming Interface" o "Interfaz de Programación de Aplicaciones". Las API son la forma en que los programas informáticos se comunican entre sí para solicitarse y enviarse información. Para esto, se le permite a una determinada aplicación de software que llame a lo que se denomina punto de conexión o end point. Twitter permite acceder a partes de su servicio mediante las API para permitir a las personas crear software que se integre con Twitter, o extraer los datos que los usuarios han decidido compartir con el mundo. Además, promueven la innovación de desarrollar programas de análisis estadísticos, análisis de sentimientos como herramientas útiles para la interacción de información personalizada (Makice, 2009).

Ilustración 5**.** Logo Python.



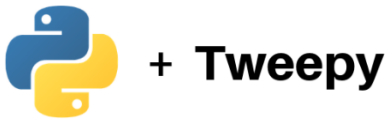
Fuente: (PYTHON, s.f.)

Python es un lenguaje de programación de código abierto de alto nivel, interactivo, siendo uno de los más populares, listaremos las ventajas por la cual optamos por este lenguaje de programación:

- Fácil de utilizar y entender, es un lenguaje interpretado por lo cual no necesita ser compilado antes de una ejecución.
- Permite realizar análisis estadístico por medio de la creación de graficas descriptivas.

- Ofrece una cantidad amplia de librerías y herramientas para tener una mayor rapidez en las tareas a realizar (Beazley, Python Essential Reference, 4th Edition, 2009).

Ilustración 6. Logo librería Tweepy.



Fuente: (Alborear, 2020)

La librería Tweepy de Python, es la interfaz oficial de aplicaciones de Twitter, que permite interactuar y manejar la información a través de la API de Twitter, la librería permite una autenticación que proporciona varios métodos de envió y entrada de los tweets, de manera directa; que va a permitir realizar el análisis exploratorio de la información, una respetiva clasificación de las fuentes que contiene el archivo y poder realizar la optimización de procesos de mejora continua para las entidades que necesitan planificar, mejorar su entorno empresarial.

El uso de la librería Tweepy es importante para la implementación del proyecto ya que permite realizar la extracción de la información de una manera ágil y simplificada de manera masiva, así mismo la librería Textblob permite realizar una limpieza y análisis de texto para determinar una polaridad en este (Alborear, 2020).

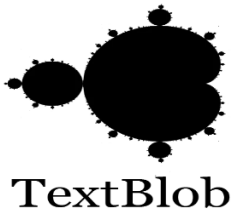
Ilustración 7. Logo Procesamiento de Lenguaje Natural.



Fuente: (Instituto de Ingenieria del conocimiento, s.f.)

El procesamiento del lenguaje natural (PLN) es un campo de la inteligencia artificial y la lingüística aplicada que se ocupa de la interacción entre computadoras y humanos mediante el lenguaje natural humano. El procesamiento del lenguaje natural tiene como objetivo desarrollar técnicas que permitan a los ordenadores entender, interpretar y generar contenido en lenguaje humano de manera efectiva, similar a como lo hace un ser humano. Esto incluye desde tareas tan básicas como analizar la gramática y el significado de una oración hasta tareas más complejas como traducción automática, resumen de textos y reconocimiento de voz. El PLN es utilizado en aplicaciones como chatbots, asistentes virtuales, motores de búsqueda y sistemas de recomendación, entre otros (Instituto de Ingenieria del conocimiento, s.f.).

Ilustración 8**.** Logo librería Textblob.



Fuente: (Ghosh & Gunning, 2019)

La librería Textblob de Python, se utiliza como procesamiento de lenguaje natural (NLP). El cual proporciona un análisis semántico, que es capaz de realizar tareas de extracciones de frases de clave, correcciones ortográficas, traducciones de idiomas, y demás tareas (Ghosh & Gunning, 2019).

Ilustración 9**.** Logo visual Studio Code



Fuente: (VISUAL STUDIO CODE, s.f.)

Visual Studio Code, es un entorno de desarrollo integrado (IDE) creado por Microsoft para Windows, Linux y macOS. Las características incluyen soporte para depuración, resaltado de sintaxis, completado inteligente de código, refactorización de código y Git incrustado, las ventajas de utilizar es IDE son:

- Admite el lenguaje de programación de Python.
- Permite la ejecución de código por partes para identificar errores de manera más rápida.
- Genera las gráficas en el mismo IDE (Ummy Gusti Salamah, Tutorial Visual Studio Code, 2021).

Ilustración 10**.** Logo librería Seaborn.



Fuente: (Waskom, 2021)

La librería Seaborn es una biblioteca de visualización de datos de Python basada en matplotlib, la cual proporciona una interfaz de alto nivel para dibujar gráficos estadísticos atractivos e informativos(Waskom, 2021). Esta librería tiene entre sus más importantes ventajas la capacidad de tener distintos tipos de visualizaciones proporcionando gráficos predeterminados

los cuales presentan un diseño bastante atractivo siendo una de las herramientas ideales para visualización de datos.

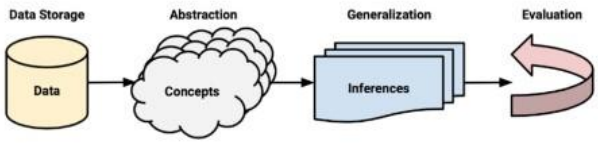
Ilustración 11**.** Logo librería WordCloud.



Fuente: (QuestionPro, s/f)

La librería WordCloud, permite crear nubes de palabras en Python; las nubes de palabras representan una de las mejores formas de analizar, ya que se encargan de mostrar visualmente la importancia de los textos de forma accesible y atractiva, resaltando las partes más interesantes de los datos en formato de texto, de esta manera las palabras que más aparecen se visualizaran en un tamaño de letra más grande y llamativa, esto hace que los análisis sean rápidos y fáciles de entender (QuestionPro, s/f).

Ilustración 12**.** Logo machine learning.



Fuente: (Lantz, 2015)

El aprendizaje automático (Machine learning), en el análisis de sentimientos, se refiere a técnicas de aprendizaje, capaz de analizar sentimientos, para poder identificar, clasificar las diferentes emociones, opiniones expresadas en los textos. Esta técnica se encargan de extraer la

información y comprender los textos de una manera positiva, negativa o neutral, donde el enfoque principal es de aprendizaje automático. Las siguientes etapas, busca como definir el aprendizaje automático:

1. Recopilar los datos que se obtienen en gran cantidad de textos.
2. Preprocesar los datos que se obtiene en los textos, para poder limpiar, eliminar palabras irrelevantes, para que el modelo obtenga un mejor proceso en el análisis de datos.
3. El entrenamiento del modelo es un algoritmo que se utiliza en aprendizaje automático, como clasificación binaria en modelos avanzados.
4. La predicción y la clasificación, consiste en el modelo que se encuentra entrenado, se pueda utilizar para nuevos textos y poder predecir la clasificación de los datos analizados (Lantz, 2015).

Ilustración 13**.** Logo aprendizaje no supervisado.



Fuente: (Inesdi Business Techschool)

El aprendizaje no supervisado es una técnica de machine learning en la que se incluyen conjuntos de datos sin etiquetar, para poder realizar análisis de datos. Con estos análisis se puede obtener información importante o vital sin poseer referencia alguna de las variables de salida. El clustering, es un método que consiste en una técnica exploratoria de análisis de datos en la que se

clasifica la información por grupos. Su objetivo principal es poder determinar la existencia de correlaciones entre las características que se presentan en los conjuntos de datos. Permitiendo de esta manera eliminar la redundancia de información y disminuyendo el tiempo de análisis (Ortega Candel, 2022).

1.1.Trabajos relacionados

En el año 2021 se realiza la investigación para la utilidad de las características lingüísticas para detectar el sentimiento de los mensajes de Twitter. El autor evalúa la utilidad de los recursos léxicos existentes, así como las funciones que capturan información sobre el lenguaje informal y creativo utilizado en microblogging. Los escritores concluyeron que es un enfoque supervisado del problema, donde se aprovecha los hashtags existentes en los datos de Twitter para crear datos de capacitación. (Moore Johanna; Kouloumpis Efthymios, 2021).

En el año 2014 se desarrolla un modelo de análisis de sentimientos, en el cual se buscó predecir los resultados de las primarias, identificando la gente que está a favor de Allamand y la gente que apoya a Longueira. De igual manera, se buscó identificar a los usuarios que están en contra de uno o ambos candidatos. Para predecir la opinión de los usuarios se diseñó un diccionario con palabras positivas y negativas con un puntaje asociado, de manera que al encontrar estos términos en los tweets se determina la polaridad del mensaje pudiendo ser positiva, neutra o negativa. El Algoritmo que se diseño tiene un acierto cercano al 60% al ocupar las 3 categorías, mientras que si sólo se ocupa para determinar mensajes positivos y negativos la precisión llega a un 74 (Montesinos García, 2014).

2. Diseño metodológico

El enfoque del presente estudio es cuantitativo, ya que principalmente emplea la recopilación de datos para su análisis y así dar respuesta a la pregunta de investigación propuesta, y basa la medición numérica y estadística para generar la propuesta de mejora.

Los datos que se utilizaron en el proceso fueron extraídos de la plataforma de Twitter con una cantidad de 1028 tuits relacionados a una organización, por motivos de confidencialidad no se relaciona el nombre, con la consulta de estos datos se logra visualizar cómo funciona el script de solución desarrollado.

2.1. Fase 1: Análisis

- Se realiza la selección de los componentes que se requieren para implementar la solución propuesta. Selección del lenguaje de programación a utilizar.
- Se realiza la búsqueda de la información de acuerdo con las necesidades de los integrantes del proyecto para adquirir el conocimiento apropiado sobre las librerías y lenguaje de programación seleccionado.
- Se elabora el diseño de la solución con el fin de mejorar el análisis de la información realizada por las entidades de bienes de consumo.

2.2. Fase 2: Diseño e integración:

- Se realiza la codificación del script propuesto según el lenguaje de programación seleccionado teniendo en cuenta los siguientes pasos:

2.2.1. Obtención de credenciales de acceso:

Se realiza la consulta de las credenciales de acceso necesarias para utilizar la API de Twitter. Estas credenciales son proporcionadas por "Twitter desarrolladores" en el momento de la creación de la cuenta.

2.2.2. Autenticación en la API de Twitter:

Por medio de la libreria tweepy en Python se realiza la autenticación a la API de Twitter para obtener acceso a las funcionalidades ofrecidas por la plataforma.

2.2.3. Recopilación de datos:

Se lleva a cabo la consulta de tuits relacionados con la entidad de bienes de consumo por medio del metodo search_tweets de la libreria tweepy. La información obtenida se almacena temporalmente para su posterior análisis.

2.2.4. Análisis de los datos:

Utilizando la librería TextBlob en Python se realiza el análisis del texto de cada tuit.

TextBlob utiliza técnicas de aprendizaje no supervisado para determinar el sentimiento percibido en el texto analizado. El resultado del análisis incluye el sentimiento percibido, el texto del tuit y la fecha en que fue publicado.

2.2.5. Representación visual de los datos:

Se utilizan las librerías Seaborn y WordCloud en Python para generar visualizaciones de los sentimientos percibidos en los tuits analizados. Estas visualizaciones ayudan a comprender mejor la percepción de los consumidores finales de manera rápida y eficaz.

2.2.6. Interpretación de resultados:

Se analizan las visualizaciones generadas y se interpretan los resultados obtenidos. Se busca identificar patrones, tendencias o insights relevantes para la entidad de bienes de consumo.

2.3. Fase 3: Implementación de las herramientas visuales:

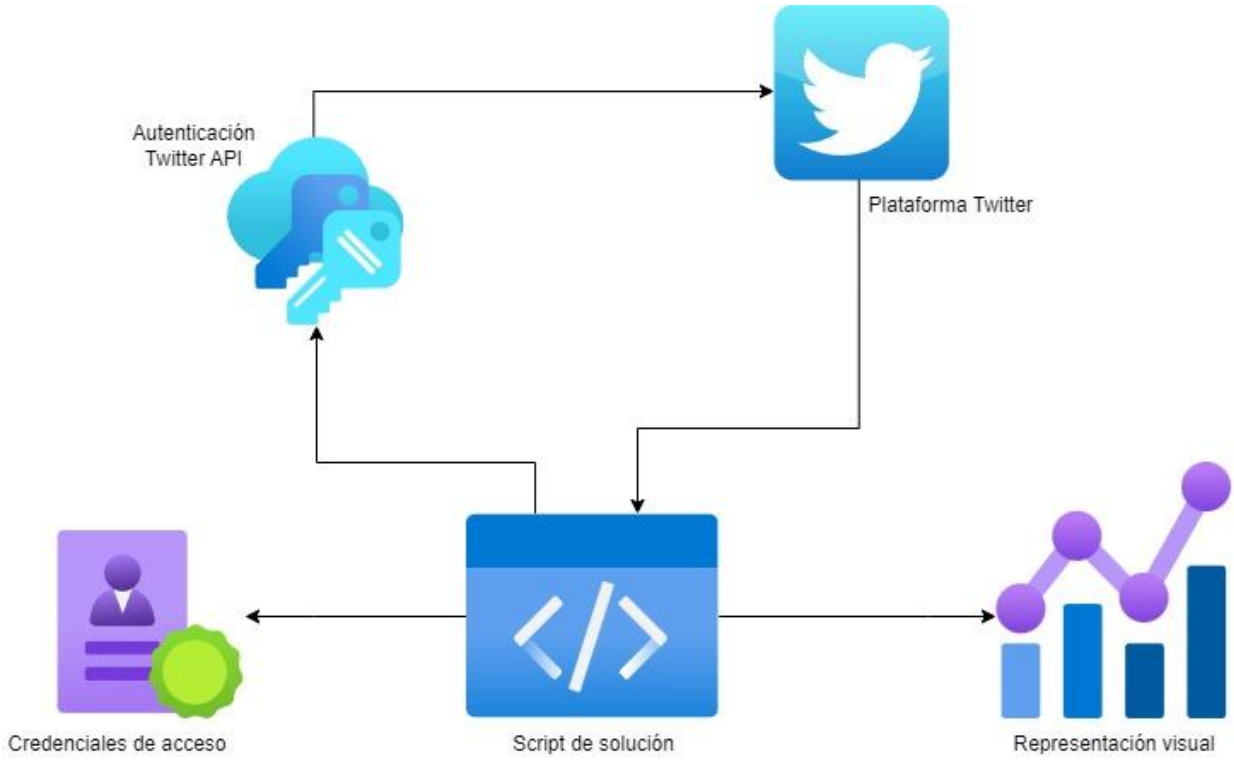
Se presentan las herramientas y posibles combinaciones generadas en el proceso a la entidad de bienes de consumo, con la información de los sentimientos percibidos de los consumidores en los datos analizados, para el uso de la entidad de bienes de consumo.

3. Resultados

3.1.Diseño de la solución.

Para llevar a cabo la solución planteada en este estudio, se diseña una solución que permite la consulta, extracción, análisis de los datos y creación de herramientas de visualización en un mismo entorno. La ilustración 12, presenta el flujo del diseño planteado.

Ilustración 14**.** Diseño de la solución.



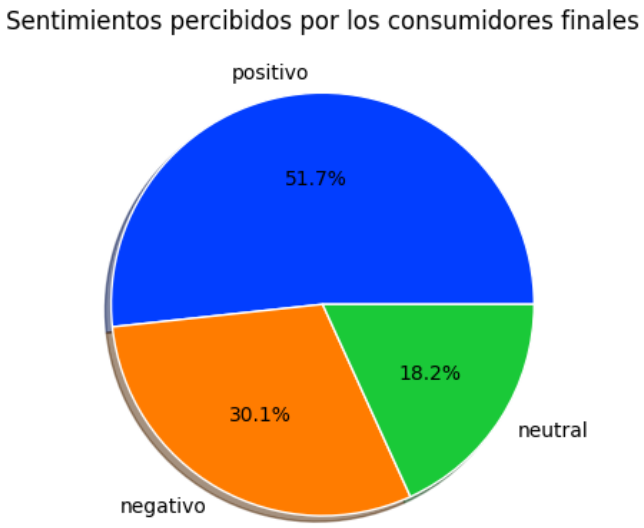
Fuente: Elaboración propia.

El script de la solución desarrollada, se entrega en un repositorio versionado, para que pueda ser implementado y ejecutado según la empresa en la que se vaya a emplear.

Repositorio: https://github.com/seminariogradotwitter/Script/tree/main

Como resultado del proceso ejecutado y con el fin de mejorar el análisis de la percepción de los consumidores finales, se generan 6 graficas; las cuales ayudaran a comprender de manera visual la percepción de las personas hacia la entidad.

Ilustración 15**.** Grafica resultado 1.



Fuente: Elaboración propia.

El uso de esta herramienta es muy útil para analizar información descriptiva, ya que permite una rápida visualización de la proporción de cada categoría dentro del conjunto de datos. Es posible comprender la relación entre las diferentes categorías y su importancia relativa en el conjunto de datos. Además, los colores y la forma circular del diagrama lo hacen reconocible y

de fácil recordación, lo que facilita la comunicación y la presentación de los resultados a otras personas.

Con la ayuda de esta herramienta grafica se logra visualizar de una manera eficaz la frecuencia relativa de los sentimientos percibidos por las personas que realizaron los tuits analizados.

Ilustración 16**.** Grafica resultado 2.

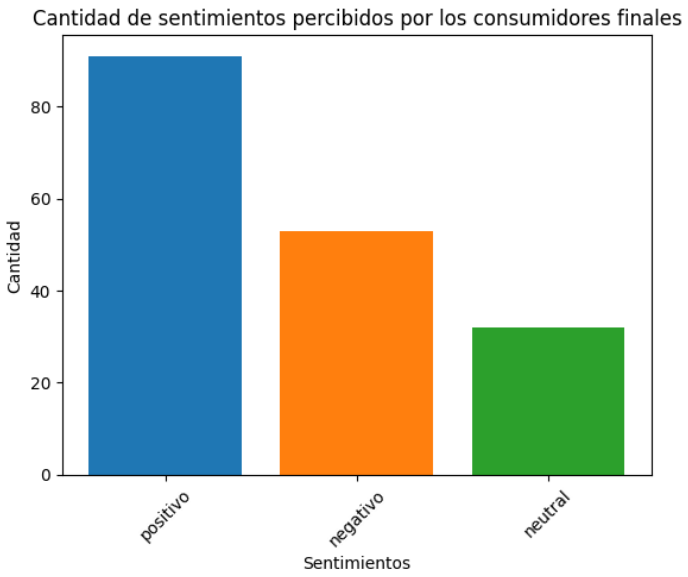


Fuente: Elaboración propia.

El uso de esta herramienta se ha vuelto cada vez más popular en el análisis de información, ya que ofrece una manera visual y concisa de resumir la información más relevante en un conjunto de datos. En lugar de tener que leer y analizar grandes cantidades de texto, la nube de palabras utiliza la frecuencia de las palabras para generar una representación gráfica de las palabras más comunes en un conjunto de datos. Esto permite a los analistas de datos identificar rápidamente los temas principales y las tendencias en la información y comprender mejor la información descriptiva de una manera visualmente atractiva y fácil de entender.

Con la ayuda de esta herramienta visual se podrá visualizar las palabras más representativas del análisis logrando identificar tendencias en las que se puede aplicar un mayor enfoque estratégico.

Ilustración 17**.** Grafica resultado 3.

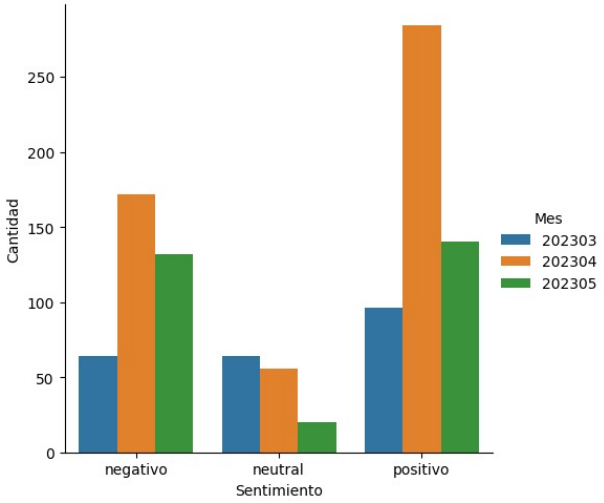


Fuente: Elaboración propia.

El diagrama de barras es una herramienta gráfica muy útil para analizar información, especialmente cuando se trata de comparar diferentes categorías o variables. Este tipo de gráfico representa valores numéricos y se utiliza comúnmente para mostrar la distribución de datos en diferentes categorías. Una de las principales ventajas de utilizar un diagrama de barras es que es fácil de entender y de comparar los diferentes valores entre sí. Otra ventaja es que los datos se pueden presentar de manera clara y concisa, lo que hace que el análisis de la información sea más fácil y rápido.

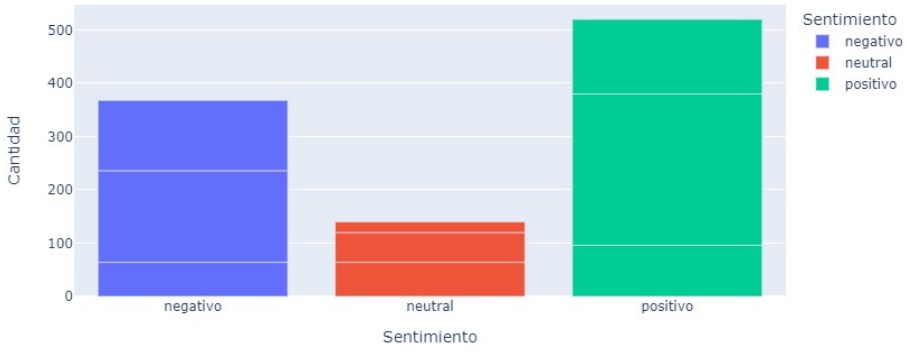
Con esta herramienta se logra visualizar la frecuencia absoluta de los sentimientos percibidos por las personas que realizaron los tuits analizados, también se logró evidenciar la cantidad de comentarios realizados en la red social Twitter.

Ilustración 18**.** Grafica resultado 4.



Fuente: Elaboración propia.

Ilustración 19**.** Grafica resultado 5.

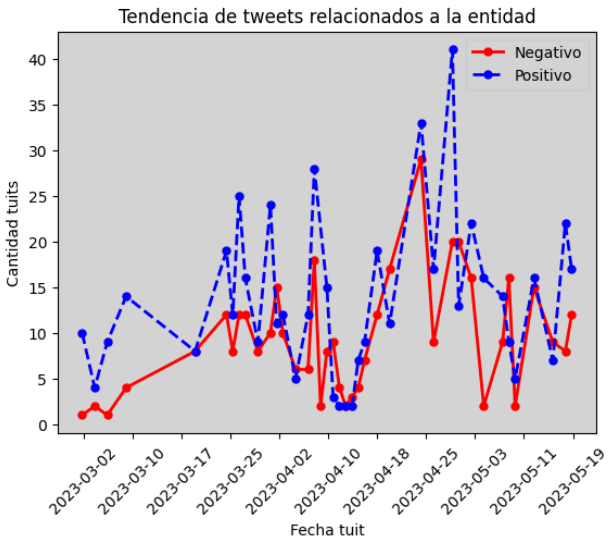


Fuente: Elaboración propia.

Con la ayuda de estas gráficas (Ilustración 18, Ilustración 19) se logra visualizar de una manera global el sentimiento percibido para cada uno de los meses, con esto se logra visualizar de una manera general los datos analizados.

Un ejemplo de interpretación que se logra con estas gráficas es el siguiente: *De acuerdo a la visualización generada se logra evidenciar una comparación de los sentimientos que se percibieron mes a mes, logrando observar que se posee una gran ventaja de los sentimientos positivos sobre los demás encontrados. Esto permite visualizar un mejoramiento continuo a través de los meses.*

Ilustración 20**.** Grafica resultado 6.



Fuente: Elaboración propia.

Con esta grafica se logra observar la cantidad de tuits en un espacio de tiempo determinado, esto ayuda a visualizar en que fechas especificas se percibió el sentimiento.

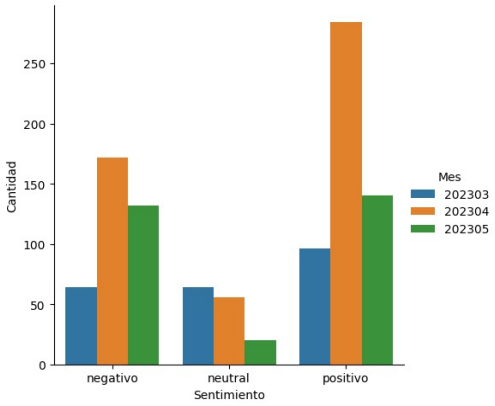
Un ejemplo de interpretación que se logra con esta gráfica es el siguiente: *De acuerdo a la visualización generada se logra evidenciar que alrededor de la fecha 20 de abril de 2023, se generó un significativo rechazo hacia la entidad, con este dato se logrará realizar una revisión*

interna y se podrá determinar el por qué se percibió este sentimiento durante dichas fechas, e implementar una posible solución.

3.2.Combinación de herramientas visuales.

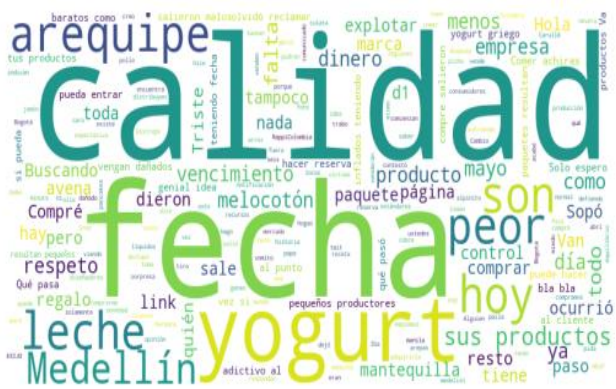
Para poder obtener una mejora potencial en el análisis de datos, se podrán combinar las herramientas visualizadas anteriormente descritas, para tener un máximo de potencial de análisis de datos.

Ilustración 22**.** Combinación análisis negativos.



Fuente: Elaboración propia.

Ilustración 21**.** Tuits Negativos



Fuente: Elaboración propia.

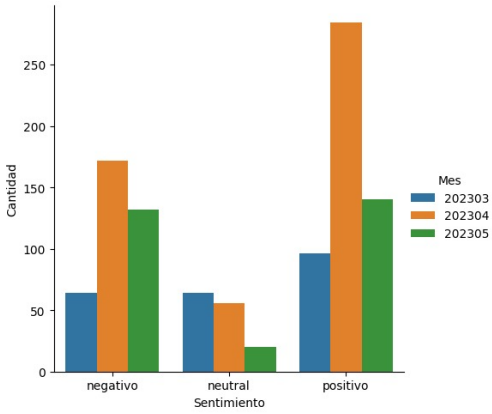
Al combinar las herramientas generadas, se logra tener una mejor visualización de las fechas y las deficiencias más notables que se obtuvieron como resultado del análisis de datos.

Un ejemplo de interpretación que se logra con la combinación de estas gráficas es el siguiente:

De acuerdo a la visualización generada se logra observar que en el mes de abril (202304) se tuvo una diferencia notaria en las líneas de calidad de los productos de la entidad

en estudio, ya que observando la gráfica de nube de palabras se ve con mayor prevalencia las palabras: "calidad", "fecha", "vencimiento", "producto"; esto ayuda a plantear planes de acción preventivos para realizar seguimiento en las líneas de producción y lograr implementar acciones de mejora que aporten a la generación productos de mejor calidad y así aumentar la satisfacción de los consumidores finales.

Ilustración 24**.** Combinación análisis positivos.



Fuente: Elaboración propia.

Ilustración 23**.** Tuits positivos



Fuente: Elaboración propia.

Al combinar las herramientas generadas, se logra tener una mejor visualización de las fechas y las fortalezas más notables que se obtuvieron como resultado del análisis de datos.

Un ejemplo de interpretación que se logra con la combinación de estas gráficas es el siguiente:

De acuerdo a la visualización generada se observa que durante el mes de mayo (202305) las campañas ejecutas obtuvieron un buen acogimiento con público. Ya que observando la gráfica de nube de palabras se ve con mayor prevalencia las palabras relacionadas a los

productos que ejecutaron una campaña durante este mes. Esto indica que los lineamientos de las campañas que se realizaron en lo trascurrido del mes de mayo pueden ser implementadas en las próximas campañas, para así generar el mismo o un mayor impacto.

3.3.Discusión De Resultados

Las herramientas visuales que se generan en el estudio pueden mejorar el análisis de la información en una entidad de bienes de consumo ya que tienen el potencial de impactar positivamente la toma de decisiones de varias maneras:

- Comprensión profunda de la percepción del consumidor: Estas herramientas permiten analizar y visualizar los sentimientos y opiniones expresados por los consumidores en Twitter. Al comprender mejor la percepción de los consumidores finales hacia la entidad y sus productos o servicios, la empresa puede identificar áreas de fortaleza y debilidad, así como detectar oportunidades de mejora.
- Detección temprana de problemas: Las herramientas visuales pueden ayudar a monitorear en tiempo real la opinión de los consumidores en Twitter. Si se identifican patrones negativos o tendencias preocupantes, la empresa puede tomar medidas rápidas para abordar problemas y evitar que se conviertan en crisis mayores. Esto permite una respuesta oportuna y eficaz a las preocupaciones de los consumidores.
- Identificación de tendencias y demandas emergentes: Al analizar los datos de Twitter, estas herramientas pueden ayudar a identificar nuevas tendencias o demandas emergentes en el mercado. Esto proporciona a la empresa información valiosa sobre las

preferencias cambiantes de los consumidores y las oportunidades para desarrollar nuevos productos, mejorar los existentes o ajustar estrategias de marketing.

- Evaluación de la efectividad de las estrategias de marketing: Las herramientas visuales permiten evaluar el impacto de las campañas de marketing en la percepción de los consumidores en tiempo real. Al analizar la reacción de los consumidores en Twitter, la empresa puede evaluar la efectividad de sus estrategias y realizar ajustes o mejoras en función de los resultados obtenidos.

las herramientas visuales que mejoran el análisis de la percepción de los consumidores en Twitter permiten a la empresa comprender mejor a su audiencia, identificar problemas y oportunidades, ajustar estrategias y mejorar la relación con los consumidores. Esto les brinda una ventaja competitiva al tomar decisiones más informadas y basadas en los datos, lo que puede impulsar el crecimiento y el éxito de la empresa en el mercado.

4. Conclusiones y recomendaciones.

4.1 Conclusiones

El análisis de la percepción del consumidor es fundamental para la gestión empresarial, y las redes sociales se han convertido en una herramienta valiosa para mejorar la capacidad de análisis de la percepción del consumidor. Las redes sociales permiten una interacción más cercana y directa con los consumidores, lo que facilita la recopilación de información valiosa sobre sus necesidades y preferencias.

Para utilizar eficazmente las redes sociales en el análisis de la percepción del consumidor, es necesario seguir las mejores prácticas en la gestión de redes sociales y contar con un equipo capacitado para analizar los datos y generar información valiosa para la empresa.

Las redes sociales se convierten en aliadas para mejorar la percepción de los clientes y usuarios hacia los productos y servicios de esta manera se pueden tomar mejores decisiones con el fin de generar los correctivos a tiempo y formular mejores estrategias.

Por medio de la percepción del consumidor sobre las redes sociales, se puede determinar que es un factor importante para poder conocer las ideas, gustos, análisis de sentimientos, donde conlleva a las empresas a mantener un margen constante de monitoreo que le va a permitir mejorar con un grado de calidad su estrategia de mercado.

Una vez que se hayan identificado las opiniones y emociones, el objetivo es que las entidades logren desarrollar estrategias para mejorar la relación con los consumidores. Esto puede incluir la implementación de políticas de atención al cliente basadas en los patrones identificados, la creación de campañas publicitarias que aborden preocupaciones específicas de los consumidores, o la introducción de nuevos productos que satisfagan las necesidades expresadas por los consumidores en los tuits.

La herramienta implementada es una ayuda importante para el análisis de sentimientos, esto debido a que está totalmente integrada aprovechando las funcionalidades que nos da Twitter para utilizar la API y las herramientas de análisis visuales generadas en Python, siendo una herramienta dinámica, amigable y fácil de poner en práctica por el usuario final.

La evaluación visual es una técnica importante en el aprendizaje no supervisado. Se pueden generar visualizaciones de los resultados para llegar a un análisis significativo y así evaluar la calidad y relevancia de los resultados obtenidos.

4.2.Recomendaciones

El presente proyecto se realizó bajo la ejecución de una licencia gratuita de Twitter desarrolladores, para la implementación de las mejoras de análisis presentadas en este proyecto se recomienda adquirir la licencia básica de Twitter desarrolladores, ya que con dicha licencia se logrará extraer un número mayor de información lo que generará mejores resultados en la ejecución del script solución desarrollado.

Dado que para generar las visualizaciones utilizadas en el proyecto se requiere un nivel de conocimiento previo de programación en Python, es recomendable el uso de otro tipo de herramientas de visualización de datos como Power BI o Tableau, ya que con el uso de estas herramientas no es necesario saber un lenguaje de programación.

Referencias

Alborear, (. P. (2020). *Research Outlook,Innovations & Research Trends in Science & Technology.*

Beazley, D. (2009). *Python Essential Reference, 4th Edition.* (Addison-Wesley, Ed.)

Borras, H. (2020). ¿Qué es un tuit?

Brunetta, H. (2013). *Marketing digital.*

Carmona, J. (2016). Analisis de sentimiento en Twitter para la predicción de precios de la bolsa. *Universidad de Valencia*.

Fan, W., Gordon, M., & Pathak, P. (2019). Sentiment analysis of Twitter data: A review of methods and applications. *Desicision Support Systems*, 1-12.

Ghosh, S., & Gunning, D. (2019). *Build Intelligent Applications that Can Interpret the Human Language to Deliver Impactful Results.* (P. Publishing, Ed.)

González-Rodríguez, M. R. (2015). *Análisis de sentimientos en Twitter para la evaluación de la imagen de marca de una empresa. El profesional de la información.*

Hennig-Thurau, T., Malthouse, E., Friege, C., Gensler, S., Lobschat, L., Rangaswamy, A., & Skiera, B. (2013). The impact of new media on customer relationships. *Journal of service research*, 371-387.

https://twitter.com/. (n.d.).

Inesdi Business Techschool. (n.d.). *https://www.inesdi.com/blog/que-es-aprendizaje-no-supervisado/.*

InfoMoney. (2020). Com cargo ameaçado, presidente do Twitter se defende.

Instituto de Ingenieria del conocimiento. (n.d.). *https://www.iic.uam.es/inteligencia/que-es-procesamiento-del-lenguaje-natural/*.

Kietzmann, J., Hermkens, K., McCarthy, I., & Silvestre, B. (2011). Social Media? Get Serius! Understanding the functional building blocks of social media. *Business Horizons*, 241-251.

Kuchinke, K. (2010). Social Media Mining y Analisis de Redes Sociales en Aplicaciones de Big Data. *Avances en Analisis y Mineria de Redes Sociales*, 679-688.

Lantz, B. (2015). *Machine Learning with R.* (P. Publishing, Ed.)

Li, X., Guo, X., & Chen, H. (2019). A segmentation method for social media audience analysis . *Journal of the Association for Information Science and Technology*, 195-208.

Lin, W. Y., Wu, C. C., & Tsai, C. Y. (2017). Consumer behavior analysis on social media for brand management. Journal of Business Research.

Makice, K. (2009). *Twitter API: Up and Running.* (O. Media, Ed.)

Montesinos García, L. (2014). *Análisis de sentimientos y predicción de eventos en twitter*. Retrieved from https://repositorio.uchile.cl/handle/2250/130479

Moore Johanna; Kouloumpis Efthymios. (2021). *https://ojs.aaai.org/index.php/ICWSM/article/view/14185/14034*.

Muñoz-Expósito, M., Oviedo-García, M. Á., & Castellanos-Verdugo, M. (2017). *How to measure engagement in Twitter: advancing a metric* (Vol. 27).

Orihuela, J. L. (2011). *Mundo Twitter.*

Ortega Candel, J. M. (2022). *Big data, machine learning y data science en python.*

PAEZ ERIKA, M. (2020). *IMPLEMENTACIÓN DE UN MODELO DE ANÁLISIS DE SENTIMIENTOS CON*. Retrieved from https://repository.ucatolica.edu.co/server/api/core/bitstreams/7f8f52aa-dc46-4030-9251-cd40b44e7bf2/content

PYTHON. (n.d.). *https://www.python.org/*.

Suarez, M. (2018). *Twitter como fuente de datos para estudios de percepción social.*

Tausczik, Y. R., & Pennebaker, J. W. (2010). *The psychological meaning of words: LIWC and computerized text analysis methods* (Vol. 29).

TWITTER. (n.d.). *https://developer.twitter.com/en*.

TWITTER. (n.d.). *https://developer.twitter.com/en/docs/twitter-api*.

Ummy Gusti Salamah, S. M. (2021). *Tutorial Visual Studio Code* (Media Sains Indonesia ed.).

VISUAL STUDIO CODE. (n.d.). *https://code.visualstudio.com/*.

Wu, S. H., Huang, S. C. T., Tsai, C. Y. D., & Lin, P. Y. (2017). Customer citizenship behavior on social networking sites: The role of relationship quality, identification, and service attributes. Internet Research, 27(2), 428–448. https://doi.org/10.1108/INTR-12-2015-0331/FULL/XML

ActiveState. (2022). What Is Pyplot In Matplotlib - ActiveState. https://www.activestate.com/resources/quick-reads/what-is-pyplot-in-matplotlib/

Alcaide, C. (n.d.). Como utilizar Twitter a fondo. Retrieved May 12, 2023, from https://www.academia.edu/en/30553350/Como_utilizar_Twitter_a_fondo

Editorial Etecé. (2021). Percepción - Concepto, etapas y componentes. https://concepto.de/percepcion/#ixzz81MmLXbWc

(PDF) Como utilizar Twitter a fondo | Carlos Alcaide - Academia.edu. (n.d.). Retrieved May 12, 2023, from https://www.academia.edu/en/30553350/Como_utilizar_Twitter_a_fondo

Platzi. (n.d.). Tutorial: cómo extraer datos de Twitter con Python. Retrieved May 12, 2023, from https://platzi.com/blog/extraer-datos-twitter/

Portal Tic. (2023). Twitter especifica las modalidades de acceso a su API, que incluye una para bots de solo lectura. https://www.europapress.es/portaltic/socialmedia/noticia-twitter-especifica-modalidad-acceso-api-incluye-bots-solo-lectura-20230330114448.html

QuestionPro. (n.d.). Nubes de palabras: Qué es, ejemplos, y pasos para crearlas. Retrieved May 12, 2023, from https://www.questionpro.com/blog/es/nubes-de-palabras/

Steven Loria. (2020). Advanced Usage: Overriding Models and the Blobber Class — TextBlob 0.16.0 documentation. https://textblob.readthedocs.io/en/dev/advanced_usage.html#advanced

Waskom, M. (2021). seaborn: statistical data visualization. Journal of Open Source Software, 6(60), 3021. https://doi.org/10.21105/JOSS.03021