



UNIVERSIDAD CÉSAR VALLEJO

FACULTAD DE INGENIERÍA Y ARQUITECTURA

**ESCUELA PROFESIONAL DE INGENIERÍA DE
SISTEMAS**

Uso del análisis de sentimientos de reseñas en restaurantes de
comidas peruanas.

TESIS PARA OBTENER EL TÍTULO PROFESIONAL DE:

Ingeniero de Sistemas

AUTOR:

Yamunaque Villegas, Henry David (orcid.org/0000-0002-5932-1103)

ASESOR:

Mg. Altuna Tocto Gerardo Arturo (orcid.org/0000-0002-8311-4788)

LÍNEA DE INVESTIGACIÓN:

Sistema de Información y Comunicaciones

LÍNEA DE RESPONSABILIDAD SOCIAL UNIVERSITARIA:

Desarrollo económico, empleo y emprendimiento

PIURA - PERÚ

2022

Dedicatoria

Dedicado a mis padres y familia por todo el apoyo y dedicación que me brindan día a día, sin ustedes nada de esto sería posible.

Agradecimiento

Agradezco a mi asesor, Magister Altuna Tocto Gerardo Arturo, por su apoyo en todo este proceso, también a mis docentes por influir de manera positiva con sus lecciones y experiencias en formarme como una persona de bien y preparada para los retos y agradezco a mi familia y amigos por que sin ellos no hubiera logrado llegar a este momento.

Índice de contenidos

Carátula	i
DEDICATORIA.....	ii
AGRADECIMIENTO.....	iii
ÍNDICE DE CONTENIDOS	iv
ÍNDICE DE TABLAS	v
ÍNDICE DE ILUSTRACIONES	vi
RESUMEN	vii
ABSTRACT	viii
I. INTRODUCCIÓN.....	1
II. MARCO TEÓRICO	4
III. METODOLOGÍA.....	12
3.1 Tipo y diseño de investigación	12
3.2. Variables y operacionalización.....	12
3.3. Población, muestra y muestreo	13
3.3.1. Población.....	13
3.3.2. Muestra.....	14
3.3.3. Muestreo.....	14
3.4. Técnicas e instrumentos de recolección de datos.....	14
3.4.1. Técnicas	14
3.4.2. Instrumentos.....	14
3.4.3. Validez.....	16
3.4.4. Confiabilidad.....	17
3.5. Procedimientos	18
3.6. Método y análisis de datos.....	19
3.7. Aspectos éticos.....	19
IV. RESULTADOS	20
V. DISCUSIÓN.....	31
VI. CONCLUSIONES.....	35
VII. RECOMENDACIONES	36
REFERENCIAS.....	37
ANEXOS	43

Índice de tablas

1: Cuadro recolección de datos I - 1	15
Tabla 2: Validez de instrumento para medir la satisfacción del usuario	17
Tabla 3: Niveles de confiabilidad de Pearson.....	18
Tabla 4: Resultados de confiabilidad de Pearson	18

Índice de ilustraciones

Ilustración 1: Análisis de sentimientos respecto al nivel – Palabras más usadas.....	21
Ilustración 2: Análisis de sentimientos respecto al nivel – Top 10 Restaurantes.	22
Ilustración 3: Análisis de sentimientos respecto a la polaridad.	23
Ilustración 4: Análisis de sentimientos respecto a la polaridad - Palabras más usadas.....	24
Ilustración 5: Análisis de sentimientos respecto a la polaridad - Top 10.	25
Ilustración 6: Análisis de sentimientos respecto a las emociones.	26
Ilustración 7: Análisis de sentimientos respecto a las emociones - Palabras más usadas.....	27
Ilustración 8: Análisis de sentimientos respecto a las emociones - Top 10.....	28
Ilustración 9: Análisis de sentimientos en reseñas de restaurantes de comidas peruanas.	29
Ilustración 10: Análisis de sentimientos en reseñas - Clusters.....	30

Resumen

Esta investigación plantea el uso del análisis de sentimientos de reseñas en restaurantes de comidas peruanas, debido a que está en auge las plataformas web y redes sociales, en las cuales se genera data, la cual con herramientas y procesos para el tratamiento de datos, se puede obtener información útil la cual ayuda a la toma de decisiones. El objetivo principal de la investigación es realizar el análisis de sentimientos de reseñas en restaurantes de comidas peruanas. Para ello se planteó una metodología de enfoque cuantitativo, de tipo aplicada, con un nivel descriptivo y diseño no experimental transeccional ya que se caracteriza por que se recolectaran datos respecto a la variable de estudio en un momento determinado. Como resultados se encontró que de las 774 reseñas procesadas, 436 son de nivel de calificación 5 en las cuales predomina la polaridad positiva, además se evidenció que el 95% de las reseñas tiene polaridad positiva, asimismo la emoción predominante es la confianza seguida de la alegría y finalmente las palabras más usadas en las reseñas son: buena, ambiente, lugar, platos y sazón. Esta investigación puede ser utilizada como base para el análisis de sentimientos en reseñas de restaurantes como en otros sectores.

Palabras clave: Análisis, sentimientos, clustering, reseñas, restaurantes.

Abstract

This research raises the use of the analysis of reviews feelings in Peruvian food restaurants, because web platforms and social networks are booming, in which data is generated, which with tools and processes for data processing, useful information can be obtained which helps decision-making. The main objective of the research is to analyze the feelings of reviews in Peruvian restaurants. For this purpose, a quantitative approach methodology was proposed, of the applied type, with a descriptive level and non-experimental transiary design since it is characterized by the collection of data regarding the study variable at a given time. As results it was found that of the 774 reviews processed, 436 are grade level 5 in which positive polarity predominates, it is also evident that 95% of the reviews have positive polarity, also the predominant emotion is confidence followed by joy and finally the most used words in the reviews are: good, atmosphere, place, dishes and seasoning. This research can be used as a basis for the analysis of feelings in restaurant reviews as in other sectors.

Keywords: Sentiment, analysis, clustering, restaurant, reviews.

I. INTRODUCCIÓN

Durante más de una década, con la llegada de Internet, es posible que los usuarios generen su propio contenido y lo compartan públicamente con mayor facilidad. En este auge, las plataformas web y de redes sociales son muy populares, lo que permite a los usuarios compartir opiniones con familiares, amigos y seguidores. Porque el objetivo principal de estas publicaciones “es expresar puntos de vista y opiniones de los usuarios” (Hernández, 2020).

Actualmente, la cantidad de datos que se generan en las plataformas web y de redes sociales es enorme (Sohangir, 2018). Esta cantidad de datos textuales, podría compararse a la basura que debería eliminarse de vez en cuando. Sin embargo, con el avance en la capacidad de almacenamiento acompañada por la creciente sofisticación en los datos, se han creado herramientas, oportunidades y desafíos para analizar y obtener información útil de esta cantidad de datos (Suarez, 2009).

Es por eso que muchas empresas aplican el análisis de sentimientos para estimar los comentarios de los clientes, publicaciones y otro tipo de contenido. Uno de los ejemplos más relevante respecto al análisis de sentimientos en la web es “el Hedonometer, un proyecto del Computational Story Lab de la Universidad de Vermont, en el cual el grupo observa más de 50 millones de tweets en inglés todos los días, para calcular una tienda de felicidad diaria” (Pozas, 2021).

“El análisis de sentimiento utiliza diferentes métodos de lingüística computacional que ayudan a detectar y extraer información subjetiva de contenidos existentes en el mundo digital como plataformas webs, redes sociales y foros ya que su principal objetivo es identificar opiniones positivas o negativas sobre productos o marcas en el texto del usuario” (Ferri y otros, 2015). Gracias al análisis de sentimiento podemos determinar si el texto extraído de Internet contiene un significado positivo o negativo en el caso más simple.

Es por eso que mediante el clustering y análisis de sentimiento, se entiende cuál es la intención exacta de una frase, saber si se refiere a una marca, a un producto en concreto o a cualquier otro aspecto.

Además el clustering “agrupa las observaciones que son similares en subconjuntos homogéneos ya que su objetivo es la detección de particiones naturales de objetos” (Guyon y otros, 2009). Estos subconjuntos pueden revelar patrones relacionados con el fenómeno en estudio, al tener estos grupos ayuda a proponer una solución particular en cada uno.

La tarea de clasificar automáticamente un texto escrito en lenguaje natural, el cual tiene emociones positivas o negativas y, a veces, es tan compleja que es difícil ponerse de acuerdo sobre la polaridad asignada a un texto dado con diferentes anotadores humanos (Zimbra, 2018). Gracias al clustering y análisis de sentimiento podemos comprender los patrones generados por los datos y de esa manera ayudar a “desarrollar mejores estrategias empresariales, facilitar la gestión de la reputación online, detectar fortalezas de la marca, aprovechar oportunidades y ayudar a la toma de decisiones para llevar a cabo en el plan estratégico de marketing” (Guerrero, 2018).

A diario se vierten cantidades de opiniones respecto al ambiente, productos, atención y precios, referentes al sector de restaurantes los cuales al ser procesados correctamente aportan claro beneficios al sector. En este contexto no cabe duda de la importancia del análisis de sentimientos de reseñas en el sector de restaurantes, ya que a través de ello se consigue clasificar automáticamente los comentarios ya sean positivos, negativos y neutros escritos por los clientes.

En relación a lo descrito anteriormente, se origina la idea de investigar ¿En qué consiste el uso del análisis de sentimientos de reseñas en restaurantes de comidas peruanas? ¿Cuál es el resultado del análisis de sentimientos en cuanto a la calificación en restaurantes de comidas peruanas? ¿Cuál es el resultado del análisis de sentimientos en cuanto a la polaridad en restaurantes de comidas peruanas? ¿Cuál es el resultado del análisis de sentimientos en cuanto a las emociones en restaurantes de comidas peruanas?

Esta investigación se justifica teóricamente dado que se desarrolla con la finalidad de incorporar al conocimiento existente del uso de análisis de sentimiento, así como el clustering, el efecto del estudio aporta conocimiento a la ciencia de la tecnología ya que va permitir que el sector de restaurantes tenga un mayor acceso a la información en un contexto visual, generando una solución para el estudio de mercado de dicho sector. El estudio se justifica de forma práctica ya que nace de solventar la necesidad de comunicar la información acerca del sector de restaurantes de manera precisa y eficiente y de esa manera contribuir a la mejora de la experiencia de los clientes.

El estudio se justifica metodológicamente ya que se utilizaron herramientas de gestión bibliográfica para almacenar y administrar las investigaciones las cuales se usaron como antecedentes. De la misma manera, se analizó y filtro cada uno de estos antecedentes, seguido de una revisión de la metodología científica utilizada en los estudios previamente mencionados, los cuales, con la ayuda de una matriz analítica, permitieron la conceptualización de las variables de estudio, entendiendo en que nivel se encuentra la ciencia sobre el tema, cómo han evolucionado las investigaciones respecto al tema en el tiempo y qué indicadores existen para admitir como referencia en esta investigación.

El objetivo general de la presente investigación es realizar el análisis de sentimientos de reseñas en restaurantes de comidas peruanas. Y se planteó como objetivos específicos determinar el análisis de sentimientos en cuanto a la calificación en restaurantes de comidas peruanas, determinar el análisis de sentimientos en cuanto a la polaridad en restaurantes de comidas peruanas, determinar el análisis de sentimientos en cuanto a las emociones en restaurantes de comidas peruanas.

II. MARCO TEÓRICO

Como resultado del análisis y selección de los materiales bibliográficos de trabajos de investigación previos, relacionados con esta investigación, se obtuvo como antecedentes el siguiente trabajo.

En primer lugar, como antecedentes internacionales está el trabajo de Cortés (2021), en su investigación del uso del análisis de sentimientos para medir el posicionamiento de una marca a través de información extraída de la red social de twitter. Su objetivo fue emplear un modelo de clasificación Naive Bayes de análisis de sentimientos para medir el posicionamiento de marca a través de la red social de twitter. Su metodología de investigación fue de tipo aplicada y de nivel descriptiva ya que describió los datos y características de la población en estudio. La metodología de desarrollo que se utilizó en la investigación se basó en la extracción de los datos mediante el api de que la red social twitter proporciona, además el algoritmo que se aplicó, está desarrollado en Python y utiliza la librería de TextBlob apoyado en nltk. Se obtuvo como resultados las polaridades de los comentarios con respecto a dichas marcas lo cual se visualizaron en nubes de palabras, además la probabilidad del tweet por palabra. Se concluyó que los valores de sentimientos pueden servir de ayuda para conocer y tratar las acciones o motivos por el cual dejaron esa opinión y quien lo hizo, además queda claro que el sentimiento puede ser manipulado por como las personas ven un producto a partir del spam.

Seguidamente está el trabajo de Osorio y Arango (2020), quienes realizaron un análisis de sentimientos mediante machine learning referente al tema de aislamiento social obligatorio causado por el Covid-19. El objetivo de esta investigación fue analizar los posibles sentimientos de las opiniones respecto al aislamiento social obligatorio en la red social Twitter, a fin de determinar los temas y palabras más resaltantes en este contexto. Su metodología de investigación fue de tipo experimental y de nivel descriptiva ya que describió los datos de la población en estudio. Se creó un algoritmo de aprendizaje automático para reconocer los sentimientos de 72.564 publicaciones y se realizó un análisis de redes sociales para identificar los temas más comunes en el

conjunto de datos. Los resultados muestran que el algoritmo tiene una alta precisión en la clasificación de sentimientos. Se determinó que el miedo fue el sentimiento principal durante todo el período de encierro en Colombia. Se concluyó que del estudio realizado se desprenden características relevantes para entender el comportamiento y los sentimientos de los ciudadanos frente al aislamiento obligatorio impuesto por las autoridades, además de identificar elementos que aportan al campo de investigación.

En el mismo ámbito se tiene a Becerra (2017) en su investigación en la Universidad Nacional de Córdoba, utilizando principalmente la red social de Twitter, realizó un estudio de la opinión de los usuarios de esta red social a partir de algunos acontecimientos que produjeron tendencia. Tuvo como objetivo de investigación proponer una solución compuesta para estudiar las opiniones que compartieron distintos usuarios de la red social Twitter. Su metodología de investigación fue de tipo aplicada y de nivel descriptiva ya que describió las características de la población. Se basó en la recolección de datos los cuales fueron pre procesados, luego se seleccionó un clasificador de sentimientos en el cual se evaluaron diferentes alternativas, por último se analizó las diferentes alternativas para visualizar la información producida durante la etapa de experimentación. Se obtuvo como resultados, después de la etapa de clasificación de sentimientos y de clustering, ver intuitivamente los temas que los usuarios comentaron en la red social de Twitter, además las palabras más utilizadas, cuál era el sentimiento asociado, la frecuencia en la que aparecen ciertos términos. Se concluyó que este proyecto estudió un problema complejo como es la minería de opiniones en la red social de Twitter y sus desafíos asociados, además se diseñó un pipeline en el cual se pre procesa, se hace clustering y representa de forma intuitiva un conjunto de tweets.

En el ámbito nacional, se encuentra Costa (2018), en su investigación sobre el uso de sistemas expertos y análisis de sentimiento en la implementación de un sistema web sobre la participación ciudadana en proyectos mineros. El objetivo de la investigación es implementar una solución técnica basada en una plataforma web para promover la participación ciudadana en el proceso de evaluación a través de la aplicación de sistemas expertos y análisis de

sentimiento, la solución permite el procesamiento dinámico y visualización de información sobre investigaciones ambientales realizadas sobre proyectos mineros. Su metodología de investigación fue de tipo aplicada y de nivel descriptiva ya que describió los datos de la población. La metodología utilizada para el desarrollo del sistema web es CommonKADS, que se basa en el modelo de ciclo de vida en espiral y nos proporciona una estructura de desarrollo del sistema web, el cual procesara todas las opiniones aportadas por los ciudadanos, se utilizaron 19200 comentarios como muestra inicial para la investigación y la verificación de los resultados. Se obtuvo como resultados del análisis del impacto del uso de la plataforma web respecto al proceso de participación ciudadana mostró que el 70% de los participantes completaron todo el proceso en el primer intento sin ayuda externa, lo que demuestra que la plataforma es intuitiva y el otro 30% de los usuarios no puede completar el proceso. La conclusión es que la implementación de una plataforma web basada en sistemas expertos y análisis de sentimiento ha mejorado el nivel de comprensión del impacto de los proyectos mineros, además, el 30% de los participantes que no completaron el proceso necesitaron capacitación para usar la plataforma.

Así mismo, Viera y otros (2019), en su investigación sobre el uso del análisis de sentimientos en Facebook respecto a la maternidad en Perú, cuyo objetivo de la investigación fue comprender las opiniones sobre la maternidad y los nuevos conceptos en el Perú. Su metodología de investigación fue de tipo aplicada y de nivel descriptiva ya que describió los datos. Se basó en la recopilación de opiniones contenidas en las publicaciones de la red social Facebook a través de una herramienta de análisis de sentimiento, que permitió analizar 28907 comentarios. Como primer resultado cuantitativo se obtuvo evidencia de opiniones positivas de la madre (5127 comentarios), lo cual tiene una gran ventaja sobre los comentarios negativos (1777). Además, la clasificación del análisis cualitativo de los comentarios condujo a la formación de 4 conjuntos de significados: la súper madre, la maternidad como fuente de bendición y de cariño, la maternidad como opción refutable y la búsqueda de nuevos conceptos de maternidad. La conclusión es que se espera que los resultados obtenidos

contribuyan a la investigación sobre la familia y la mujer, además, a través de este análisis, se han hallado conceptos que se espera que se continúen investigando, pero no representa el concepto de población peruana, pues si bien la muestra es amplia, se ha seleccionado aleatoriamente.

Del mismo modo, Alva (2021), en su investigación de análisis del sentimiento político en twitter durante las elecciones congresales 2020 en el Perú. El objetivo de la investigación es desarrollar un sistema para analizar los sentimientos políticos de los ciudadanos peruanos, utilizando como corpus un conjunto de tweets obtenidos de Twitter durante las elecciones Congresales 2020 en Perú. Su metodología de investigación fue de tipo experimental y de nivel descriptiva ya que describió los datos población en estudio. Para el desarrollo de esta herramienta se adoptó un método híbrido combinado con el método CRISP-DM. Los resultados muestran que las mejores herramientas de pago relacionadas con el análisis de sentimiento son Searcher y Brand 24, además identificar bibliotecas de código abierto como Scikit-Learn y Pyspark, también se utilizó datos extraídos de Twitter los cuales se limpiaron utilizando diferentes técnicas de pre procesamiento, también se identificaron 4 modelos de aprendizaje supervisado y se desarrolló un sistema basado en la tecnología MEAN. Se concluyó que esta investigación ha desarrollado con éxito un sistema para analizar el sentimiento político en Twitter aplicable a las elecciones al Congreso de 2020.

Seguidamente, se describen los fundamentos teóricos de la presente investigación, para el estudio de la variable análisis de sentimiento se detalla la siguiente teoría:

En primer lugar se menciona al Análisis de sentimiento, de tal término se presenta la siguiente definición, para Aguirre (2018), el análisis de sentimientos es un campo de estudio que tiene como objetivo utilizar algoritmos para extraer automáticamente opiniones y sentimientos sobre las entidades y sus aspectos del lenguaje natural del texto. Además para Amigo (2020), el análisis de sentimientos nos permite identificar y obtener información subjetiva de una opinión. También para Pérez y Amparo (2017), el análisis de sentimientos se

centra en determinar la polaridad general de un texto es decir determinar el efecto emocional. Uno de los métodos de análisis de sentimientos se basa en utilizar algún léxico, los cuales clasifican el sentimiento en función de las polaridades de palabras predefinidas. (Zunic y otros, 2020).

Uno de los diccionarios de léxico más utilizado es del NRC, este diccionario de léxico contiene aproximadamente 27.000 palabras y se basa en el léxico de afectos del Consejo Nacional de Investigación de Canadá (NRC) y los conjuntos de sinónimos de WordNet de la biblioteca NLTK. De esta forma, se puede analizar el texto y obtener resultados, como emociones o sentimientos (Isasi 2021).

Para el estudio de la variable “Reseñas en restaurantes de comidas peruanas” se detalla la siguiente teoría:

Según Lang (2021), los restaurantes son un comercio que se caracteriza por brindar alimentos y bebidas para los clientes, además suele consistir en un amplio lugar en el cual se puede disfrutar de un ambiente acogedor y se puede socializar. En los restaurantes los clientes dejan reseñas, según SocialwiBox (2020), las reseñas en restaurantes de comidas peruanas son opiniones sinceras de los usuarios respecto al negocio. Son ellos quienes dan información valiosa a otros usuarios que buscan un restaurante. Es por eso que en esta investigación se toma en cuenta los siguientes conceptos. En cuanto a la calificación, según Martínez (s/f), es una expresión compuesta (numérica o conceptual) de una evaluación. Según la Rae (s/f), los sentimientos son el estado emocional de la mente o hecho o efecto de sentir o sentirse. Según la Rae (s/f), la polaridad son los rasgos semánticos de elementos sintácticos que requieren contexto positivo o negativo. Según Blanco (2019), las emociones es el conjunto de cambios fisiológicos, cognitivos y motores que surgen de la evaluación consciente o inconsciente de estímulos, en un contexto dado y en relación con las metas de un individuo en un momento dado de su vida.

En cuanto a la metodología de desarrollo, para realizar el prototipo y proceso de análisis de sentimientos se implementaran una serie de pasos donde cada uno

dependerá del anterior, los cuales son: recolección de reseñas, pre procesamiento, análisis de sentimientos, clustering y visualización.

Para la etapa de recolección de datos (reseñas) se utilizara Web scraping en las plataformas de TripAdvisor y GoogleMaps. Web scraping, es una técnica que extrae y almacena datos de una o más páginas web para analizarlos o manipularlos en otros medios, además las herramientas de web scraping nos ayuda extraer datos de sitios web, de forma rápida, eficiente y de manera automatizada, ofreciendo datos en un formato más estructurado y más fácil de usar (Castrillo, 2015).

En la etapa de pre procesamiento de datos, las reseñas se someterán a una serie de técnicas que permitirán limpiar y reducir las características de los textos, así facilitando el análisis de sentimientos: Filtrado, se utiliza para eliminar palabras innecesarias, como puntuación, emojis, números, espacios, etc. En este trabajo se eliminaron las preposiciones y conjunciones consideradas palabras vacías y emojis (Allahyari, 2017). Tokenización: El objetivo es dividir oraciones en palabras o frases y generar fragmentos más pequeños llamados tokens. En este estudio, las reseñas se dividirán por caracteres (palabras), y luego las etiquetas o tokens se utilizaran en otras etapas de procesamiento (Demidova & Klyueva, 2017). Eliminación de stop words: Se utiliza para eliminar palabras innecesarias y sin sentido, como conjunciones y artículos (por ejemplo: y, desde, hasta, etc.). El propósito de este paso es limpiar el texto, dejando solo las palabras relevantes (Pratama, 2019).

En la etapa de clustering con los datos ya previamente procesados realizaremos el proceso de agrupar y separar las opiniones de los usuarios utilizando el algoritmo K-Means. Según Rubio y Alba (2019), el clustering es una técnica de reducción de datos que se utiliza específicamente para interpretar y encontrar subgrupos de observaciones en un conjunto de datos, estos subgrupos se obtienen por el alto grado de similitud de objetos. Además Según Mamani (2015), el clustering o agrupación divide la información en diferentes grupos. El propósito de la agrupación es encontrar grupos (clúster) muy diferentes a otros, y sus miembros son muy similares entre sí. Además el método de agrupamiento

intenta establecer grupos de objetos en un espacio de n-dimensiones, de modo que los objetos que están más cerca quedan agrupados en el mismo grupo y que los distintos grupos estén lejanos entre sí. Además nos menciona que el algoritmo k-means tiene como objetivo minimizar la suma de las distancias entre los puntos y centroide al que pertenecen, además el funcionamiento paso a paso de k-means es el siguiente, primero se realiza la elección del número de clusters lo cual vendría a ser K, luego se inicializa las coordenadas(aleatorias) de los centroides, así mismo se asigna a cada punto un clúster, para lograr esto, primero se calcula la distancia y cada punto a cada centroide para luego agruparlas con aquel centroide más próximo, por consiguiente se recalculan los centroides de los clusters y luego se repiten los últimos pasos hasta llegar al criterio de parada. Además según Becerra (2017), para poder saber que tan efectivas son las predicciones del clasificador se debe evaluar el rendimiento mediante una validación cruzada de K iteraciones. Cuando hablamos de rendimiento de una clasificador estamos haciendo referencia al accuracy del mismo, la proporción de los casos que fueron clasificados adecuadamente sobre el total de todos los casos a clasificar.

$$accuracy = \frac{Total\ de\ aciertos}{Total\ de\ casos}$$

En la etapa de análisis de sentimientos, realizaremos el análisis de sentimientos utilizando un clasificador, el cual etiquetara el sentimiento expresado por cada reseña y de esa manera tener una apreciación global por el sentimiento expresado.

En la etapa de visualización, los resultados obtenidos al realizar el análisis de sentimiento se mostraran de manera intuitiva utilizando diferentes gráficos para el fácil entendimiento por parte de los usuarios. La visualización de datos es el proceso de adquirir, interpretar y comparar datos para comunicar claramente

ideas complejas, facilitando así la identificación y análisis de patrones significativos. (Bartlett, s/f).

Según ISO 90001 (2015), un KPI es un nivel cuantificable del logro de un objetivo crítico, derivado directamente de, o por medio de una función de agregación de mediciones físicas, datos u otros KPIs.

Además para el desarrollo del prototipo y proceso se hará uso el lenguaje de Python y R. Según Marzal y Gracia (2009), Python es un lenguaje de programación interpretado, de múltiples paradigmas y multiplataforma, que se utiliza principalmente para Big data, IA (inteligencia artificial), ciencia de datos, marcos de prueba y desarrollo web, se caracteriza por su simplicidad, velocidad, curva de aprendizaje corta y amabilidad. Está desarrollado bajo una licencia de código abierto, por lo que se puede utilizar y distribuir de forma gratuita. Según Santana y Farfán (2014), R es un entorno de software libre (licencia GNU GPL) y lenguaje de programación interpretado. El término ambiente pretende caracterizarlo en un sistema completamente planificado y coherente, en lugar de una acumulación de herramientas específicas e inflexibles, que a menudo ocurre en otro software de análisis de datos.

III. METODOLOGÍA

3.1 Tipo y diseño de investigación

Tipo: La presente investigación es de tipo aplicada porque según el autor Schwarz (2017) es el empleo de los conocimientos en la práctica ya que se centra en resolver un problema en provecho de la sociedad.

Diseño: La presente investigación es No – experimental Transeccional, ya que se recolectarán datos respecto a la variable de estudio en un momento o tiempo determinado.

Nivel: La presente investigación tiene un nivel de investigación descriptiva y propositiva, ya que permite describir los datos y características del fenómeno o población en estudio. (Marroquín, s/f).

Enfoque: La presente investigación es cuantitativa porque según el autor Schwarz (2017) se centra en medir variables y magnitudes de los aspectos del problema.

3.2. Variables y operacionalización

En la presente investigación se establecieron dos variables de estudio, la primera es, análisis de sentimiento.

- Definición Conceptual: Análisis de sentimiento es definido como el conjunto de técnicas informáticas que se utilizan para extraer, clasificar, comprender y evaluar opiniones expresadas en fuentes publicadas en Internet, comentarios en portales y otro contenido generado por el usuario.(Cambria & Hussain, 2012).
- Definición Operacional: Es una variable que será construida a fin de mejorar la experiencia de los clientes. Las dimensión de la presente variable es: modelo.
- Indicadores: En los indicadores de la variable se define medir la precisión del modelo de análisis de sentimientos.
- Escala de medición: La escala de medición para todos los indicadores será ordinal.

Del mismo modo la segunda variable de estudio, la cual ha sido denominada “Reseñas en restaurantes de comidas peruanas”.

- Definición Conceptual: Las reseñas son opiniones sinceras de los usuarios de nuestro negocio. Son ellos quienes dan información valiosa a otros usuarios que buscan un restaurante. (SocialwiBox, 2020).
- Definición Operacional: Es una variable que será evaluada con el fin de determinar el nivel de cada indicador. Las dimensiones de la presente variable son: calificación, polaridad y emociones.
- Indicadores: En los indicadores de la variable se definió medir el nivel de calificación, polaridad positiva, polaridad negativa, alegría, asco, anticipación, confianza, enfado, miedo, sorpresa y tristeza.
- Escala de medición: La escala de medición para los indicadores serán nominales.

3.3. Población, muestra y muestreo

3.3.1. Población

Para el presente estudio la población estará constituida por las reseñas de restaurantes de comida peruana en un rango de los últimos 3 meses.

Según Marroquín (s/f) la población de estudio es un conjunto de datos que pertenecen al ámbito espacial donde se desarrolla el trabajo de investigación.

- Criterios de inclusión: Reseñas de restaurantes de solo comida peruana, solamente de los últimos 3 meses.
- Criterios de exclusión: Reseñas de restaurantes de comidas de otros países, reseñas que tengan más de 3 meses de antigüedad y reseñas que contenga emojis.

3.3.2. Muestra

La muestra del presente estudio está conformada por toda la población de dicha investigación por tal motivo no existe una muestra como tal.

Según Marroquín (s/f) la muestra es un fragmento representativo de la población en la que se llevará a cabo la investigación.

3.3.3. Muestreo

El muestreo para dicha investigación será omitida ya que la muestra está tomando toda la población de estudio.

Según López (2004) el muestreo es un método utilizado para seleccionar partes constituyentes de una muestra de una población.

3.4. Técnicas e instrumentos de recolección de datos

3.4.1. Técnicas

Se usará como técnica de recolección de datos, el análisis documental porque según el autor Liniers (s/f) es un conjunto de operaciones mediante el cual extraemos nociones del documento para representarlo con la finalidad de posibilitar su recuperación posterior o identificación.

3.4.2. Instrumentos

Para el análisis y estudio de las dimensiones e indicadores referentes a los grupos de elementos en reseñas de restaurantes de comidas peruanas, se usará como instrumento de evaluación la ficha de registro.

Según Hernández (2014) el instrumento de recolección de datos es un recurso importante que utilizan los investigadores, para registrar información o datos sobre las variables de estudio.

Según Serrano (2018), las fichas de registro son herramientas o instrumentos prediseñados en las que se diseñan los aspectos observables para facilitar la observación estructurada.

Tabla 1: Cuadro recolección de datos I - 1

N.º	Indicador	Técnica	Instrumento
1	Nivel	Análisis documental	Ficha de registro.
2	Polaridad Positiva		
3	Polaridad Negativa		
4	Alegría		
5	Asco		
6	Anticipación		
7	Confianza		
8	Enfado		
9	Miedo		
10	Sorpresa		
11	Tristeza		

Fuente: Elaboración propia

3.4.3. Validez

Se define como el grado en que las pruebas realizadas en el instrumento nos proporcionan información adecuada y coherente. (Villasís, 2018).

Validez de constructo: Se refiere al grado en que un instrumento de medición cumple con los supuestos esperados de un instrumento de medición diseñado para medir con precisión lo que se pretende medir. Puede pensarse como un concepto general que abarca otros tipos de validez. (Chiner, s/f).

Validez de contenido: Se refiere a la medida en que la prueba proporciona una muestra adecuada de lo que cubre, sin omisiones ni contenido desigual. (Chiner, s/f).

Validez de criterio: Se refiere al grado en que la prueba se relaciona con una variable (criterio) fuera de la prueba y la expectativa de que debería estar relacionada de alguna manera. (Chiner, s/f).

En el presente estudio se aplicara el criterio de experticia de los jueces al instrumento, el cual es una ficha de registro que servirá para evaluar cuantitativamente los indicadores propuestos.

Además el criterio de experticia es una de las técnicas utilizadas para calcular el índice de validez estructural. Se basa en la correspondencia teórica entre los conceptos de ítems instrumentales y eventos. Intenta confirmar el consenso entre investigadores y expertos de que cada proyecto pertenece a la respectiva sinergia de eventos. (Soriano, 2015).

Tabla 2: Validez de instrumento para medir la satisfacción del usuario

N°	Experto	Grado académico	Puntaje
1	Correa Calle, Teófilo Roberto	Magister en Ing. de Sistemas	94
2	Jaramillo Atoche, Javier Eduardo	Magister en Ing. de Sistemas	95
3	Altuna Tocto, Gerardo Arturo	Magister en Ing. de Sistemas	72.5
	Promedio		87.2

Fuente: Elaboración propia

Como se puede observar en la tabla N°2, respecto al instrumento de medición de las dimensiones de calificación, polaridad y emociones, se envió la respectiva tabla de validación, la cual fue validada por los expertos tal cual se observa en los anexos N°4, N°5 y N°6. Se obtuvo un promedio de 87.2% de nivel de confianza, entonces el instrumento cumple para la medición de las dimensiones planteadas.

3.4.4. Confiabilidad

La confiabilidad se define como el grado de consistencia de las puntuaciones obtenidas por un mismo grupo de sujetos en una serie de medidas sobre el mismo instrumento. (Villasís, 2018).

Técnica: Se aplicara la prueba de test–retest en dos tiempos diferentes para evaluar la confiabilidad de los instrumentos. La prueba de test-retest consiste en realizar la misma prueba en la misma muestra de sujetos en dos ocasiones diferentes con un intervalo de tiempo determinado. Además calcular el coeficiente de

correlación entre las puntuaciones obtenidas por los sujetos en estas dos situaciones. (Chiner, s/f).

Tabla 3: Niveles de confiabilidad de Pearson

Escala	Nivel
$0.00 < \text{sig.} < 0.20$	Muy bajo
$0.20 \leq \text{sig.} < 0.40$	Bajo
$0.40 \leq \text{sig.} < 0.60$	Regular
$0.60 \leq \text{sig.} < 0.80$	Aceptable
$0.80 \leq \text{sig.} < 1.00$	Elevado

Fuente: Elaboración propia

Tabla 4: Resultados de confiabilidad de Pearson

	Resultado	Porcentaje
Pearson	0.99	99%

Fuente: Elaboración propia

Como se puede observar en la tabla N°4, el resultado que se obtiene de la prueba test retest de confiabilidad es de un 0.99 es decir un 99% lo cual indica que el instrumento cumple con lo necesario para la medición de las dimensiones planteadas en este estudio.

3.5. Procedimientos

En relación a la presente investigación, se utilizaron las fichas de registro para el ingreso de los datos de las reseñas de los restaurantes de comidas peruanas las cuales fueron extraídas mediante las técnica de web scraping de la fuente google maps, donde la extracción se basó en los datos más relevantes para el estudio, además aquellos datos se sometieron a un pre

procesamiento en el cual se logró limpiar y reducir características o palabras innecesarias como puntuación, números, espacios, emojis y otros, es decir los textos se seleccionaron para evitar la inconsistencia de la data y así facilitar el análisis de sentimientos.

3.6. Método y análisis de datos

Una vez finalizada la recolección de datos, se procedió a la etapa de análisis de los datos de acuerdo a las dimensiones e indicadores de cada variable. En este caso específico del análisis de sentimientos, en el procesamiento de los datos, se realizó mediante la aplicación de técnicas de agrupamiento (clustering) y el proceso de análisis de sentimientos donde se obtuvo un conjunto de resultados. Como la investigación fue carácter descriptiva, se hizo uso de la visualización de datos mediante diagramas de barra, gráficos y otros.

Además el análisis estadístico de minería de datos como gráficos de nube de palabra, gráfico de dispersión. Para ello se utilizó el software estadístico R.

3.7. Aspectos éticos

Esta investigación se basa en el Código de Ética en Investigación de la Universidad César Vallejo, que tiene como objetivo promover la integridad científica de la investigación en el ámbito de la Universidad César Vallejo, respetar los más altos estándares de rigor científico, rendición de cuentas e integridad y garantizar la autenticidad del conocimiento científico protege los derechos de investigación y el bienestar de los participantes, investigadores y los derechos de propiedad intelectual.

El investigador es el único responsable de realizar la investigación planificada de forma estricta y exhaustiva para prevenir cualquier tipo de error en los resultados. Además, se dará a conocer un informe completo, cuyo propósito absoluto es no alterar los datos recuperados y eliminar la cercanía que obstaculiza la objetividad y corrección del trabajo.

IV. RESULTADOS

Luego de haber extraído los datos (reseñas) de los restaurantes de Piura mediante la técnica de web scraping, además haber realizado el pre procesamiento de los datos en los cuales se aplicaron las técnicas de filtrado, tokenización y eliminación de stop words, además haber realizado la etapa de análisis de sentimientos y clustering se obtuvieron los siguientes resultados.

OE01: Determinar el análisis de sentimientos en cuanto a la calificación en restaurantes de comidas peruanas.

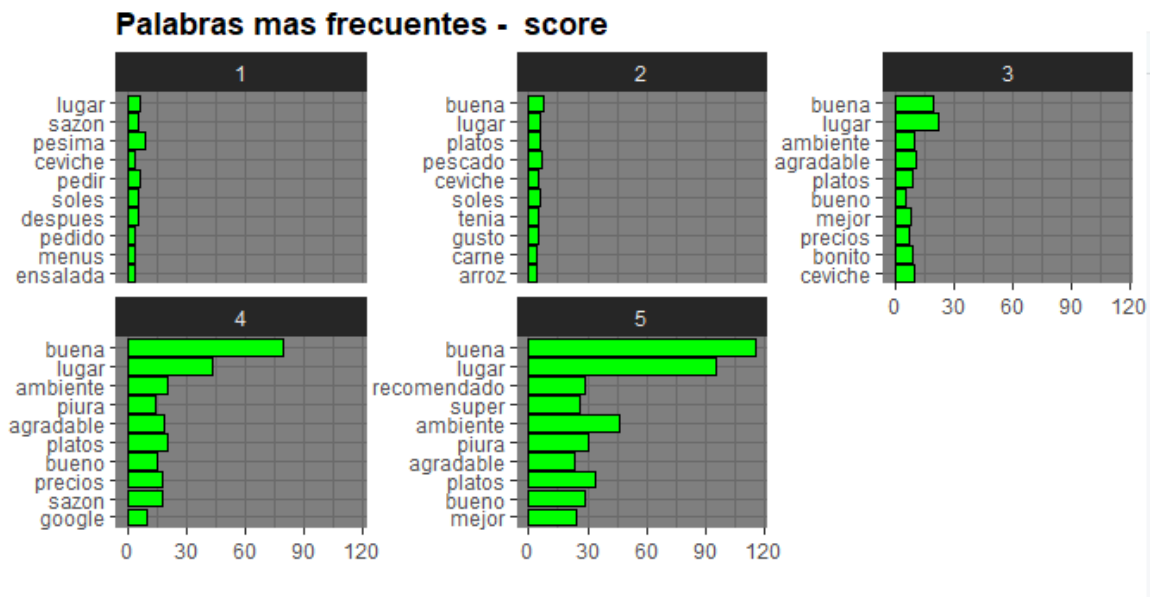
Tabla 5: *Resultados de nivel de calificación*

Nivel	N° Reseñas
1	35
2	32
3	79
4	192
5	436

Fuente: Elaboración propia

Para la dimensión calificación, que tiene como indicador el nivel de calificación, para ello, se usaron 774 reseñas de 103 restaurantes. De acuerdo a la estructura de la datos recopilados mediante la técnica web scraping se obtuvo como muestra la tabla 5, que 35 reseñas son de nivel 1, 32 reseñas tiene son de nivel 2, 79 reseñas son de nivel 3, 192 reseñas son de nivel 4 y 436 reseñas son de nivel 5.

Ilustración 1: Análisis de sentimientos respecto al nivel – Palabras más usadas.



Fuente: Elaboración propia – R Studio

Como se presenta en la ilustración 1 las 5 palabras más usadas dentro de las reseñas de *nivel 1* son: *lugar*, *pésima*, *pedido*, *carne* y *ensalada*, asimismo en el *nivel 2* son: *cantidad*, *expectativa*, *carne*, *ensalada* y *cobrar*, además en el *nivel 3* son: *lugar*, *platos*, *agradable*, *ceviche* y *ambiente*, también en el *nivel 4* son: *buena*, *lugar*, *Piura*, *platos* y *ambiente*, por último en el *nivel 5* son: *buena*, *lugar*, *ambiente*, *carta* y *agradable*. Por lo tanto las palabras más usadas dentro de los diferentes niveles son: *ambiente*, *lugar*, *buena*, *carta* y *platos*, además se menciona *precios*, *música*, *ceviche* y *sazón*.

Ilustración 2: Análisis de sentimientos respecto al nivel – Top 10 Restaurantes.

Restaurante	Nivel / N° Reseñas					Total
	5	4	3	2	1	
L'Ancora Grill & Lounge	17	9	3	2	1	32
Casa Andina Premium Piura	25	5	2	-	-	32
Restaurante Julia	23	7	-	-	-	30
Filomeno Brasa Peruana	15	8	2	-	1	26
El Uruguayo	17	5	1	1	1	25
Restaurant La Tomasita	12	8	1	2	2	25
El Nuevo Ajicito	9	10	3	-	-	22
D'resaca (picanteria)	12	6	1	3	-	22
LongHorn Piura	9	5	3	-	4	21
La Pera Madura	11	6	-	2	1	20

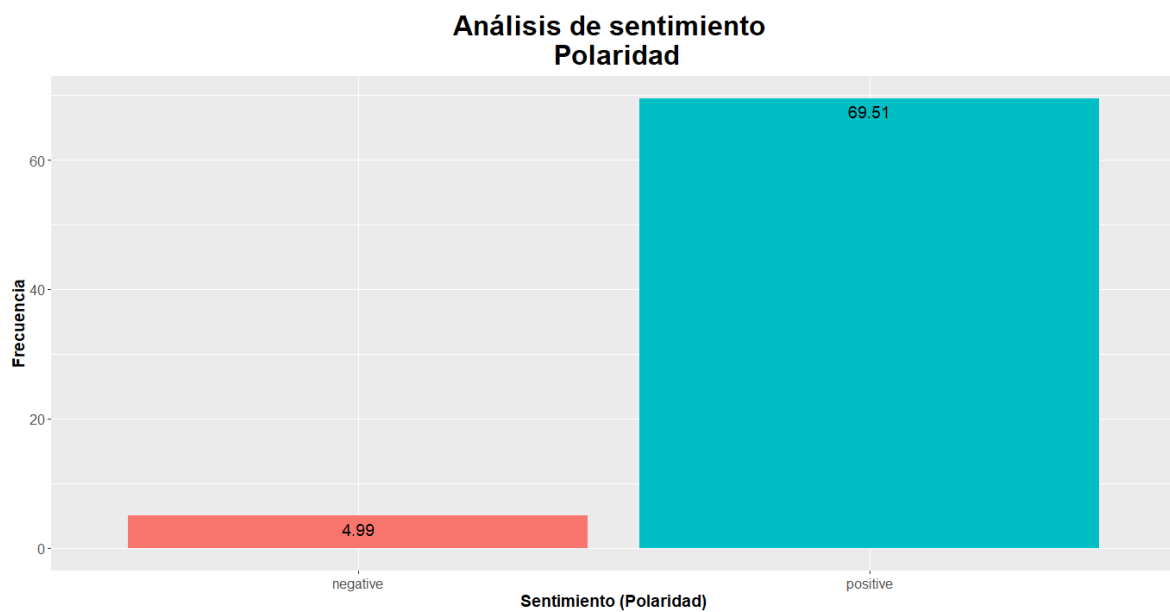
Fuente: Elaboración propia - Google Data Studio

Como se presenta en la ilustración 2 se puede observar que en el top de los 10 restaurantes con más reseñas, predomina el nivel 5 en calificación, seguido del nivel 4 y que al menos cada restaurante cuenta con una reseña en nivel 1.

OE02: Determinar el análisis de sentimientos en cuanto a la polaridad en restaurantes de comidas peruanas.

En los indicadores de polaridad, positivos y negativos, para los resultados evaluados se obtuvo que en la polaridad negativa se obtiene una frecuencia o puntuación de 4.99 y para los sentimientos positivos se obtiene una frecuencia o puntuación de 69.51.

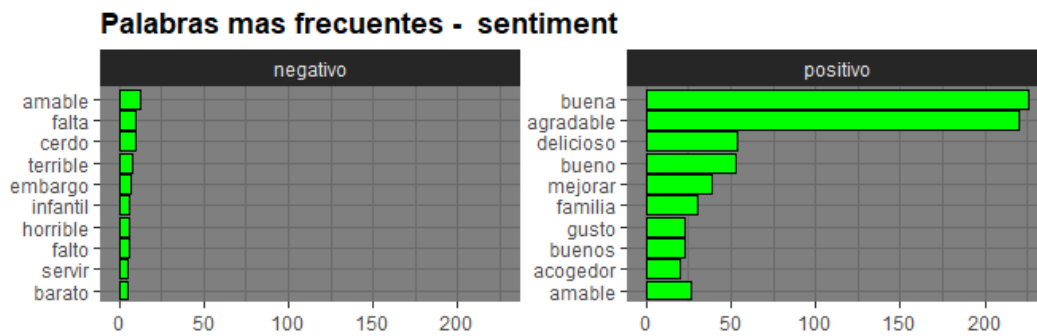
Ilustración 3: *Análisis de sentimientos respecto a la polaridad.*



Fuente: Elaboración propia - R Studio

Como se presenta en la ilustración 3, se obtiene una frecuencia o puntaje positivo de 69.51 en las 774 reseñas analizadas y un 4.99 de frecuencia o puntaje negativos en las 774 reseñas, lo cual nos da entender que la mayoría de las reseñas son positivas.

Ilustración 4: Análisis de sentimientos respecto a la polaridad - Palabras más usadas.



Fuente: Elaboración propia - R Studio

Como se presenta en la ilustración 4, las 5 palabras más usadas dentro de las reseñas según la polaridad **positiva** (opinión que es una afirmación positiva) son: buena, agradable, delicioso, familia y mejorar y también para la polaridad **negativa** (opinión que es una afirmación negativa) son: amable, falta, cerdo, terrible e infantil.

Ilustración 5: *Análisis de sentimientos respecto a la polaridad - Top 10.*

	Restaurante	positive	negative
1.	Beef House - La Casa de Maco	1,4	0,08
2.	Restaurant Cevichería El Ajicito	0,66	0,19
3.	EL ENSUEÑO - RESTAURANT	0	0
4.	Rocovan	1	0
5.	Restaurant Cevicheria El Ollero	0,38	0,13
6.	La Burguesia Piura	0,33	0
7.	La Isla Del Piuranito	1	0
8.	De Mar Y Fuego	0,86	0
9.	El Nuevo Ajicito	0,81	0,01
10.	Restaurant Sol Y Arena	0,71	0,04

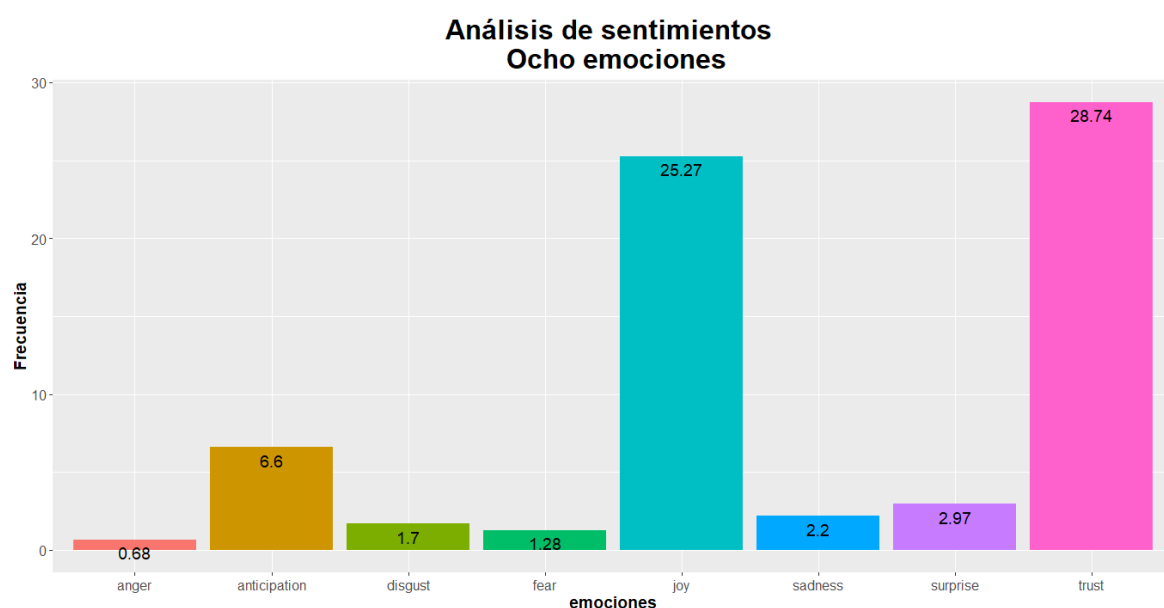
Fuente: Elaboración propia - Google Data Studio

Como se presenta en la ilustración 5, se puede observar que en el top de los 10 restaurantes según la polaridad, el restaurante “Beef House” tiene la polaridad positiva más alta con 1.4 asimismo reseñas con una polaridad negativa de 0.08. Además se puede observar que 5 de los restaurantes tiene una polaridad negativa 0, lo que implica que no posee ningún comentario negativo dentro de las reseñas analizadas.

OE03: Determinar el análisis de sentimientos en cuanto a las emociones en restaurantes de comidas peruanas.

La dimensión de emociones en donde se encuentran 8 indicadores lo cuales viene a ser la emociones de alegría, asco, anticipación, confianza, enfado, miedo, sorpresa y tristeza. Para ello se evaluó mediante el proceso y modelo del análisis de sentimiento. Una vez realizado dicha evaluación y proceso, resulto que la emoción con mayor puntaje o frecuencia es confianza y que la emoción con menor puntaje o frecuencia es enfado.

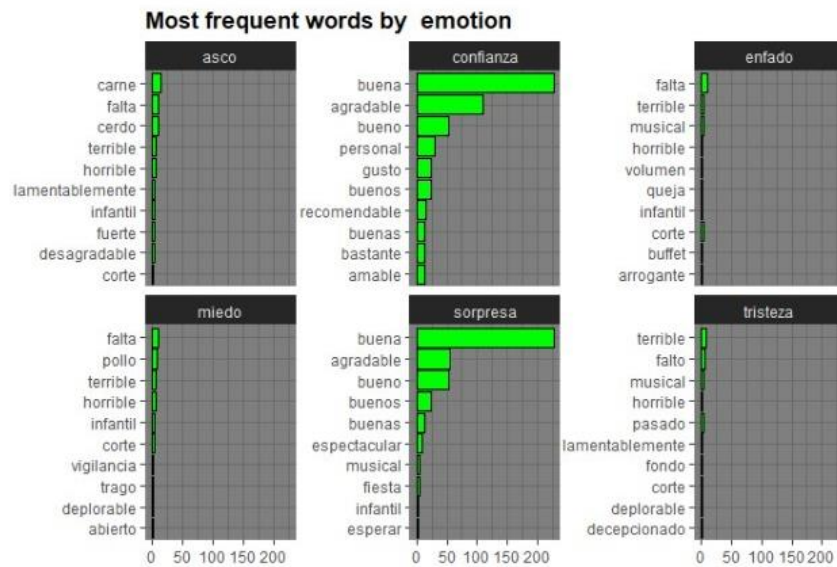
Ilustración 6: *Análisis de sentimientos respecto a las emociones.*



Fuente: Elaboración propia R – Studio

Como se presenta en la ilustración 6, la emoción de alegría tiene una frecuencia de un 25.27, asco 1.7, anticipación 6.6, confianza 28.74, enfado 0.68, miedo 1.28, sorpresa 2.97 y por ultimo tristeza un valor de 2.2. La frecuencia se interpreta como la cantidad de veces que hace referencia en las reseñas a cada uno de estas 8 emociones más representativas.

Ilustración 7: Análisis de sentimientos respecto a las emociones - Palabras más usadas.



Fuente: Elaboración propia - R Studio

Como se presenta en la ilustración 7 las 5 palabras más usadas dentro de las reseñas según la emoción predominante confianza son: buena, agradable, personal, gusto y recomendable. Asimismo en la emoción sorpresa se muestra que dentro de las palabras más usadas esta: espectacular y música.

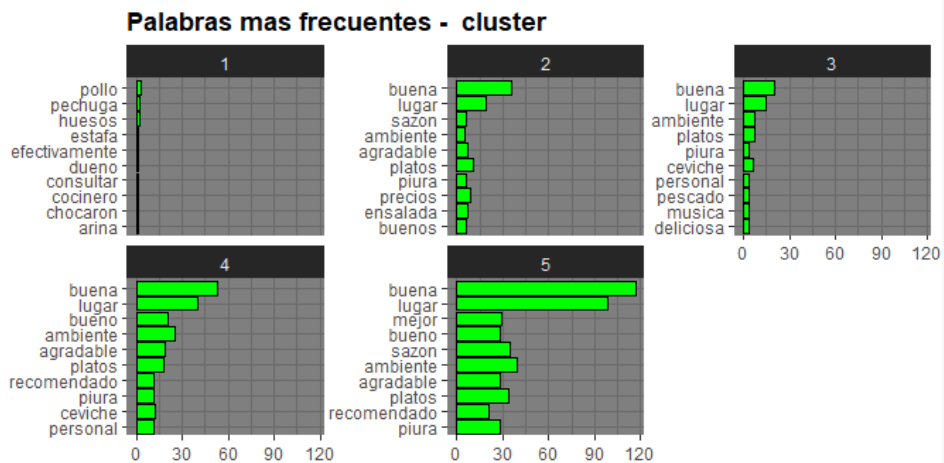
Ilustración 8: Análisis de sentimientos respecto a las emociones - Top 10.

Restaurante	anger	anticipation	disgust	fear	joy	sadness	surprise	trust
1. Restaurant Cevicheria Don Belly	0	0	0	0	0,33	0	0	0,67
2. La Habana	0	0	0	0	0,4	0	0	0,6
3. Mr. Leigh	0	0,1	0	0	0,37	0	0	0,53
4. Bebé Mero - Restaurant	0	0	0	0,03	0,34	0,08	0,03	0,51
5. Caffè Ristorante Da Biagio	0	0	0	0	0,5	0	0	0,5
6. Pirata's	0	0	0	0	0,5	0	0	0,5
7. Lahaina	0	0	0	0	0	0	0	0,5
8. Sabor Piurano	0	0	0	0,5	0	0	0	0,5
9. La Fonda, Castilla - Piura	0	0	0	0	0,5	0	0	0,5
10. Restaurant Doña Bárbara	0	0	0	0	0,5	0	0	0,5

Fuente: Elaboración propia Google Data Studio

Como se presenta en la ilustración 8, se puede observar que en el top de los 10 restaurantes según las emociones, el restaurante “Don Belly”, tiene el más alto puntaje de 0.67 relacionado con la emoción de confianza, seguido de “La Habana” con una puntuación de 0.6, pero a la vez se visualiza que ambos restaurantes, poseen un puntaje alto relacionado con la emoción alegría.

Ilustración 10: Análisis de sentimientos en reseñas - Clusters.



Fuente: Elaboración propia - R Studio

Como se presenta en la ilustración 10, mediante el proceso de clustering se clasifico en 5 clusters según el puntaje del análisis de sentimientos, lo cual muestra las palabras más usadas en la reseñas según cada cluster.

V. DISCUSIÓN

En la presente investigación, se realizó el uso del análisis de sentimientos, basándose en 3 dimensiones, calificación, sentimientos y emociones con respecto a “reseñas de restaurantes de comida peruana”.

Se evaluó la variable “reseñas en restaurantes de comidas peruanas” respecto a la dimensión de calificación, el cual tuvo como objetivo determinar el análisis de sentimientos en cuanto a la calificación en restaurantes de comidas peruanas en donde se pudo evidenciar una diferencia media entre los niveles de calificación, como se evidencia en la tabla 5 e ilustración 1.

Estos resultados son equivalentes con los resultados de Cortés (2021), quien en su estudio hizo un análisis de sentimientos para medir el posicionamiento de una marca a través de información extraída de la red social de twitter utilizando un modelo clasificador y herramienta de análisis como lo es Naive Bayes y VADER, se pudo obtener que la marca que contiene una puntuación mayor se encuentra con mejor calificación por el público, además las marcas que tienen mejor calificación son las que tienen mayor tweets positivos, lo cual en los resultados de mi investigación se refleja exactamente en la ilustración 1, que las reseñas de restaurantes que tienen mayor porcentaje de sentimiento positivo son las que tienen una mejor calificación, además al igual que el top de marcas, en la presente investigación también se realizó un top de restaurantes tomando en cuenta el nivel de calificación por reseña para su segmentación, el cual se puede ver en la ilustración 2, asimismo se realizó un gráfico de las palabras más usadas por nivel de calificación en vez de una nube de palabras, en el cual se pudo obtener que las palabras más usadas son ambiente, lugar, buena y carta, esto se puede evidenciar en la ilustración 1.

Asimismo, en esta investigación se obtuvo que las reseñas analizadas, la mayoría tiene un puntaje positivo respecto a la polaridad y fueron mínimas las negativas lo cual se observa en la ilustración 3, también se hizo una segmentación específica obteniendo un top 10 respecto a los restaurantes en el cual se obtiene la polaridad positiva y negativa por restaurante, donde en cada restaurante predomina la polaridad positiva, aunque en muchos de estos restaurantes no

cuentan con reseñas que tengan un puntaje negativo respecto a la polaridad, esto también se puede evidenciar en la ilustración 5, finalmente en la ilustración 4 en la cual se hace un gráfico por las palabras más usadas según la polaridad, cabe resaltar que la polaridad positiva es la que predomina ante la polaridad negativa.

En algunas investigaciones como la de Osorio y Arango (2020), quienes en su estudio realizaron un análisis de sentimientos mediante machine learning referente al tema de aislamiento social obligatorio causado por el Covid-19, se pudo obtener una de las emociones predominante en todo este análisis de sentimientos con respecto al Covid-19 fue el miedo lo cual con la ampliación del tiempo de la cuarentena la emoción de la tristeza lo supero. Tomando en cuenta esto en el análisis de sentimiento en reseñas de restaurantes de comidas peruanas, la emoción que predomina es la confianza a la cual le sigue la alegría lo cual se puede evidenciar en la ilustración 6, además en la ilustración 8 se realizó un top con respecto a restaurantes en cual también se evidencia por restaurante que la emoción predominante es la confianza, además la ilustración 8 tuvo como objetivo hacer que este resultado respecto a los emociones sea sencillo entender y lograr obtener conclusiones específicas, cabe resaltar que también se realizó un gráfico de las palabras más usadas por cada emoción para su mejor visualización y comprensión de los resultados en el cual también se observa claramente que la emoción predominante es la confianza, según la ilustración 7.

Asimismo en la misma investigación de Osorio y Arango (2020), se centró en la comparación de 3 algoritmos para el análisis de sentimientos Modelo Deep Learning, Naive Bayes y Naive Bayes Kernel, tomando como base los comentarios respecto a Covid-19, además en la investigación de Alva (2021) se mostró que las mejores herramientas de pago relacionadas con el análisis de sentimiento son Searcher y Brand 24, también se identificó bibliotecas de código abierto como Scikit-Learn y Pyspark tomando como base los datos extraídos de Twitter, mientras tanto en la actual investigación se usó un modelo clasificador con respecto a un diccionario de léxico el cual contiene aproximadamente 27.000 palabras y se basa en el léxico de afectos del Consejo Nacional de Investigación

de Canadá (NRC), además como base se tomaron las reseñas en restaurantes de comida peruana.

También en la misma investigación de Osorio y Arango (2020), se observó que usaron una metodología de 7 etapas, selección de palabras clave, extracción de información, procesamiento del texto, entrenamiento del modelo, evaluación de algoritmos de ML, análisis de sentimientos y de redes y por último resultados, mientras tanto en esta investigación la metodología usada se basó en 5 pasos, extracción de la información, procesamiento del texto, análisis de sentimientos, clustering y resultados, se usó esta metodología por que no contó con la comparación de algoritmos y menos con la selección de palabras clave ya que la fase de selección se tomó en cuenta en la fase de la extracción de la información como algo opcional ya que se tenía definido el sector (reseñas de restaurantes) y filtros para la data que se necesitaba obtener y evaluar.

También en la investigación de Becerra (2017) en la cual utilizó principalmente la red social de Twitter, realizó un estudio de la opinión de los usuarios de esta red social a partir de algunos acontecimientos que produjeron tendencia en el cual se tomó en cuenta la frecuencia del uso de palabras, las cuales la mayoría estaban en tendencia, además se aplicó el proceso de clustering y se clasificó de manera gradual entre los términos y palabras con significado negativo a positivo. Respecto a esto, la actual investigación sobre análisis de sentimientos en reseñas de restaurantes de comidas peruanas, obtuvo cuales son las palabras más usadas en las diferentes dimensiones e indicadores, además las reseñas se clasificaron en 5 clúster a los cuales se le realizó un gráfico con las palabras más usadas lo cual se puede observar en la ilustración 10.

Asimismo, en la investigación de Becerra (2017) se utilizó el gráfico de nube de palabras para la visualización de los datos por cada clúster lo cual en esta investigación se logró realizar de manera general la cual se puede observar en la ilustración 9, cabe resaltar que las palabras más usadas en todo el análisis de sentimientos en reseñas de restaurantes de comidas peruanas son buena, lugar y platos, además se realizó el gráfico de barras por cada palabra más usada dependiendo de cada dimensión e indicador, la misma que se puede observar en las ilustraciones 1, 4, 7 y 10.

Por último, mi investigación ha querido indagar la percepción actual respecto a los restaurantes, para poder identificar la imagen de aquellos sobre todo en Piura, es por eso que como fuente se utilizó las reseñas de la plataforma google maps mediante la técnica de análisis de sentimientos, el objetivo fue hacer uso del análisis de sentimientos en las reseñas de los restaurantes de comidas peruanas, obteniendo como resultados que la mayoría de las reseñas son positivas y además se obtuvo que las palabras lugar, buena, platos, ambiente y agradable son las más usadas y esto se puede evidenciar en la ilustración 9, pero cabe resaltar que estas palabras depende mucho del contexto en que los comentarios fueron realizados, es decir la palabra guarda diversos significados, tal cual se menciona en la investigación de Viera y otros (2019), es por eso que también deberían hallarse categorías al agruparse los significados de acuerdo a la importancia de la semántica dentro de la palabra, además el análisis de sentimientos realizado se puede implementar en un sistema web como se menciona en la investigación de Costa (2018), lo cual ayudaría a tener un procesamiento dinámico y visualización de la información para beneficio del sector.

VI. CONCLUSIONES

A partir de los resultados de esta investigación se obtuvieron las siguientes conclusiones:

1. En lo que respecta al uso del análisis de sentimientos en cuanto a calificación en reseñas de restaurantes de comidas peruanas, se logró evidenciar que de las 774 reseñas, 436 reseñas tienen una calificación 5 las cuales tienen un sentimiento positivo.
2. Para el uso del análisis de sentimientos en cuanto a las emociones (alegría, asco, anticipación, confianza, enfado, miedo, sorpresa y tristeza), en reseñas de restaurantes de comidas peruanas se puede evidenciar que la emoción de la confianza es la emoción predominante seguida de la emoción alegría y anticipación.
3. En lo que respecta al uso del análisis de sentimientos en cuanto a la polaridad en reseñas de restaurantes de comidas peruanas se logró evidenciar una frecuencia o puntaje positivo de 69.51 en las 774 reseñas y un 4.99 en frecuencia o puntaje negativo en las 774 reseñas.
4. El objetivo principal del presente estudio, fue realizar el análisis de sentimientos de reseñas en restaurantes de comidas peruanas, se logró evidenciar que la mayoría de reseñas son positivas en las cuales el sentimiento predominante es la confianza, además que las palabras más usadas en las reseñas son: buena, ambiente, lugar, platos, sazón y otros.

VII. RECOMENDACIONES

A partir de los resultados de esta investigación se desprenden diferentes recomendaciones que se pueden trabajar e implementar:

1. Se podría tener una mayor participación y contribución en la parte del pre procesamiento de las reseñas, el cual debería ser más agresivo para que las características obtenidas permitan tener un puntaje más efectivo y preciso en el análisis de sentimientos y clustering.
2. Asimismo se podría implementar una serie de etapas más complejas para que se contemplen diferentes particularidades del lenguaje natural, como también tomar en cuenta ciertos aspectos referentes a los restaurantes.
3. Además se podría tener en cuenta la mejora en la visualización de datos, se recomienda encontrar unos gráficos o esquema que permita procesar visualmente los datos de una manera más rápida y efectiva ya que resulta algo complejo por la serie de pasos y procesamiento de los datos.
4. Asimismo tomar en cuenta que las tecnologías usadas en esta investigación aun requieren de un arduo trabajo de búsqueda y experimentación para ser implementados, allí la necesidad de profundizar los conocimientos y explorar nuevas soluciones que nos brinden resultados más precisos.
5. Finalmente, esta investigación puede ser utilizada como base para el análisis de sentimientos en reseñas de restaurantes u otros sectores, gracias a los resultados, en donde con el análisis de sentimientos se logró obtener información importante para la toma de decisiones.

REFERENCIAS

1. BECERRA, M., 2017. Análisis de sentimientos en Twitter: el bueno, el malo y el >:(. *XX Concurso de Trabajos Estudiantiles - JAIIO 46 (Córdoba, 2017)*. [en línea]. S.l.: s.n., [Consulta: 20 noviembre 2021]. Disponible en: <http://sedici.unlp.edu.ar/handle/10915/65440>.
2. Reseñas de restaurantes, ¿por qué son importantes y cómo conseguirlas? *Software Wifi Social* [en línea], 2020. [Consulta: 7 noviembre 2021]. Disponible en: <https://www.socialwibox.es/resenas-de-restaurantes-conseguirlas/>.
3. ALVA SEGURA, D.A., 2021. Análisis del Sentimiento Político en Twitter durante las Elecciones Congresales 2020 en el Perú. En: Accepted: 2021-05-19T09:53:33Z [en línea], [Consulta: 7 noviembre 2021]. Disponible en: <https://reunir.unir.net/handle/123456789/11364>.
4. MAMANI RODRÍGUEZ, Z.E., 2015. Aplicación de la minería de datos distribuida usando algoritmo de clustering k-means para mejorar la calidad de servicios de las organizaciones modernas caso: Poder judicial. [en línea]. [Consulta: 7 noviembre 2021]. Disponible en: <https://cybertesis.unmsm.edu.pe/handle/20.500.12672/4472>.
5. ARANGO PASTRANA, C.A y OSORIO ANDRADE, C.F., 2020. Aislamiento social obligatorio: un análisis de sentimientos mediante machine learning – *Revista Suma de Negocios*. [en línea]. [Consulta: 7 noviembre 2021]. Disponible en: <http://revistasumadenegocios.konradlorenz.edu.co/vol12-num-26-2021-aislamiento-social-obligatorio-un-analisis-de-sentimientos-mediante-machine-learning/>.
6. VIERA, R.A.S., CERNADES, L.P.C., OCHOA-LUNA, J.E., BASAURI, A.P.T.-V. y CÁCERES, C.M.D., 2019. Maternidad en Perú a través del uso del Sentiment Analysis en Facebook. *Revista Latina de Comunicación Social*, no. 74, pp. 1031-1055. ISSN 1138-5820. DOI [10.4185/RLCS-2019-1370](https://doi.org/10.4185/RLCS-2019-1370).
7. AGUIRRE RIVADENEIRA, C., 2018. Estudio del análisis de sentimiento en redes sociales para la prescripción de situaciones financieras. En:

Accepted: 2018-11-13T12:41:23Z [en línea], [Consulta: 20 noviembre 2021]. Disponible en: <https://ruidera.uclm.es/xmlui/handle/10578/19059>.

8. URIBE, J.F.C., 2018. Análisis de algoritmos de clustering para datos categóricos. [en línea], Disponible en: <https://repositorio.uniandes.edu.co/bitstream/handle/1992/39125/u820979.pdf?sequence=1>.
9. CAMBRIA, E., GRASSI, M., HUSSAIN, A. y HAVASI, C., 2012. Sentic Computing for social media marketing. Multimedia Tools and Applications, vol. 59, no. 2, pp. 557-577. ISSN 1573-7721. DOI [10.1007/s11042-011-0815-0](https://doi.org/10.1007/s11042-011-0815-0).
10. LÓPEZ, P.L., 2004. POBLACIÓN MUESTRA Y MUESTREO. Punto Cero, vol. 09, no. 08, pp. 69-74. ISSN 1815-0276.
11. HERNÁNDEZ, C., 2020. Conoce la importancia que tienen los comentarios en redes sociales • Bright Marketing. Bright Marketing [en línea]. [Consulta: 12 noviembre 2021]. Disponible en: <https://www.brightmarketing.mx/conoce-la-importancia-que-tienen-los-comentarios-en-redes-sociales/>.
12. POZAS, J.L.B., 2021. ¿Qué es el análisis de sentimientos? Usando PNL y ML para extraer significado. CIO MX [en línea]. [Consulta: 12 noviembre 2021]. Disponible en: <https://cio.com.mx/que-es-el-analisis-de-sentimientos-usando-pnl-y-ml-para-extraer-significado/>.
13. GUERRERO NIETO, M., 2018. Análisis del sentimiento en Turismo en redes sociales - IIC. Instituto de Ingeniería del Conocimiento [en línea]. [Consulta: 12 noviembre 2021]. Disponible en: <https://www.iic.uam.es/digital/analisis-del-sentimiento-turismo-redes-sociales/>.
14. GUYON, I., LUXBURG, U.V. y WILLIAMSON, R.C., 2009. Clustering: Science or art. NIPS 2009 Workshop on Clustering Theory [en línea]. S.I.: s.n., Disponible en: <http://proceedings.mlr.press/v27/luxburg12a/luxburg12a.pdf>.
15. FERRI, F., D'ANDREA, A., GRIFONI, P. y GUZZO, T., 2015. Approaches, Tools and Applications for Sentiment Analysis Implementation. International

- Journal of Computer Applications, vol. 125, pp. 26-33. DOI [10.5120/ijca2015905866](https://doi.org/10.5120/ijca2015905866).
16. ISO 9001, 2015. ¿Qué son los indicadores KPI y qué tipos existen? - ISO 9001:2015. [en línea]. [Consulta: 13 noviembre 2021]. Disponible en: <https://www.nueva-iso-9001-2015.com/2019/10/que-son-los-indicadores-kpi-y-que-tipos-existen/>.
 17. LANG, G., 2021. Restaurant | Description, History, Types, & Purpose | Britannica. [en línea]. [Consulta: 13 noviembre 2021]. Disponible en: <https://www.britannica.com/topic/restaurant>.
 18. ASALE, R.- y RAE, [sin fecha]. sentimiento | Diccionario de la lengua española. «Diccionario de la lengua española» - Edición del Tricentenario [en línea]. [Consulta: 27 septiembre 2022]. Disponible en: <https://dle.rae.es/sentimiento>.
 19. ASALE, R.- y RAE, [sin fecha]. polaridad | Diccionario de la lengua española. «Diccionario de la lengua española» - Edición del Tricentenario [en línea]. [Consulta: 28 septiembre 2022]. Disponible en: <https://dle.rae.es/polaridad>.
 20. MARTINEZ DEL CAMPO, L., 2016. Más allá de la calificación. Instrumentos para evaluar el aprendizaje, Concepción: Universidad de Concepción, 2016. S.l.: s.n. ISBN 978-956-9280-18-4.
 21. BLANCO CANALES, A., 2019. la emoción y sus componentes, pp. 7.
 22. VILLASÍS-KEEVER, M.Á., MÁRQUEZ-GONZÁLEZ, H., ZURITA-CRUZ, J.N., MIRANDA-NOVALES, M.G. y ESCAMILLA-NÚÑEZ, A., 2018. El protocolo de investigación VII. Validez y confiabilidad de las mediciones. Revista Alergia México, vol. 65, no. 4, pp. 414-421. ISSN 2448-9190, 0002-5151. DOI [10.29262/ram.v65i4.560](https://doi.org/10.29262/ram.v65i4.560).
 23. SORIANO RODRÍGUEZ, A.M., 2015. Diseño y validación de instrumentos de medición. Diá-logos, no. 14, pp. 19-40. ISSN 1996-1642. DOI [10.5377/dialogos.v0i14.2202](https://doi.org/10.5377/dialogos.v0i14.2202).
 24. CASTRILLO, O., 2015. Web Scraping: Applications and Tools. , no. 2015, pp. 31.
 25. ALLAHYARI, M., POURIYEH, S., ASSEFI, M., SAFAEI, S., TRIPPE, E.D., GUTIERREZ, J.B. y KOCHUT, K., 2017. A Brief Survey of Text Mining:

- Classification, Clustering and Extraction Techniques. En: arXiv: 1707.02919, arXiv:1707.02919 [cs] [en línea], [Consulta: 16 noviembre 2021]. Disponible en: <http://arxiv.org/abs/1707.02919>.
26. DEMIDOVA, L. y KLYUEVA, I., 2017. Improving the Classification Quality of the SVM Classifier for the Imbalanced Datasets on the Base of Ideas the SMOTE Algorithm. ITM Web of Conferences, vol. 10, pp. 02002. ISSN 2271-2097. DOI [10.1051/itmconf/20171002002](https://doi.org/10.1051/itmconf/20171002002).
27. PRATAMA, B.T., UTAMI, E. y SUNYOTO, A., 2019. An optimization of a Lexicon Based Sentiment Analysis Method on Indonesian App Review. 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE). S.l.: s.n., pp. 341-346. DOI [10.1109/ICITISEE48480.2019.9003900](https://doi.org/10.1109/ICITISEE48480.2019.9003900).
28. BARTLETT, R., [sin fecha]. Introduction to Data Visualization. , pp. 25.
29. COSTA BOJANICH, B.N., 2018. Implementación de plataforma web de participación ciudadana para evaluación de proyectos mineros mediante sistema experto y análisis de sentimientos. En: Accepted: 2019-07-23T16:09:43Z, Universidad Nacional Mayor de San Marcos [en línea], [Consulta: 20 noviembre 2021]. Disponible en: <https://cybertesis.unmsm.edu.pe/handle/20.500.12672/10650>.
30. CORTÉS DÍAZ, N.S., 2021. Uso del análisis de sentimientos para medir el posicionamiento de una marca a través de la información extraída de la plataforma Twitter. En: Accepted: 2021-09-24T12:36:34Z [en línea], [Consulta: 20 noviembre 2021]. Disponible en: <https://repository.usta.edu.co/handle/11634/35771>.
31. ZIMBRA, D., ABBASI, A., ZENG, D. y CHEN, H., 2018. The State-of-the-Art in Twitter Sentiment Analysis: A Review and Benchmark Evaluation. ACM Transactions on Management Information Systems, vol. 9, no. 2, pp. 5:1-5:29. ISSN 2158-656X. DOI [10.1145/3185045](https://doi.org/10.1145/3185045).
32. SOHANGIR, S., WANG, D., POMERANETS, A. y KHOSHGOFTAAR, T.M., 2018. Big Data: Deep Learning for financial sentiment analysis. Journal of Big Data, vol. 5, no. 1, pp. 3. ISSN 2196-1115. DOI [10.1186/s40537-017-0111-6](https://doi.org/10.1186/s40537-017-0111-6).

33. SUÁREZ, Yuniét Rodríguez; AMADOR, Anolandy Díaz. Herramientas de minería de datos. Revista Cubana de Ciencias Informáticas, 2009, vol. 3, no 3-4, p. 73-80.
34. ZUNIC, A., CORCORAN, P. y SPASIC, I., 2020. Sentiment Analysis in Health and Well-Being: Systematic Review. En: Company: JMIR Medical Informatics Distributor: JMIR Medical Informatics Institution: JMIR Medical Informatics Label: JMIR Medical Informatics publisher: JMIR Publications Inc., Toronto, Canada, JMIR Medical Informatics, vol. 8, no. 1, pp. e16023. DOI [10.2196/16023](https://doi.org/10.2196/16023).
35. ISASI, J., 2021. Análisis de sentimientos en R con «syuzhet». Programming Historian [en línea], [Consulta: 20 noviembre 2021]. Disponible en: <https://programminghistorian.org/es/lecciones/analisis-de-sentimientos-r>.
36. AMIGO PORTILLA, A., 2020. Análisis de sentimiento en Twitter. Reputación de las principales aerolíneas europeas tras la crisis sanitaria del covid-19. En: Accepted: 2021-05-06T07:23:46Z [en línea], [Consulta: 20 noviembre 2021]. Disponible en: <https://dspace.unia.es/handle/10334/5915>.
37. PÉREZ, E. y AMPARO, M., 2017. El impacto de las emociones en el análisis de la polaridad en textos con lenguaje figurado en Twitter. En: Accepted: 2017-09-01T06:29:35Z [en línea], [Consulta: 20 noviembre 2021]. Disponible en: <https://riunet.upv.es/handle/10251/86127>.
38. RUBIO TAPIAS, J.A. y ALBA RODRÍGUEZ, C.A., 2019. Desarrollo de componente web parametrizable de clustering para análisis de datos. En: Accepted: 2019-10-08T19:51:03Z [en línea], [Consulta: 20 noviembre 2021]. Disponible en: <https://repository.ucatolica.edu.co/handle/10983/23778>.
39. SANTANA, J.S. y FARFÁN, E.M., 2014. El arte de programar en R: un lenguaje para la estadística. , pp. 197.
40. MARZAL VARÓ, A. y GRACIA LUENGO, I., 2009. Introducción a la programación con Python [en línea]. S.I.: Universitat Jaume I. [Consulta: 20 noviembre 2021]. ISBN 978-84-692-5869-9. Disponible en: <http://repositori.uji.es/xmlui/handle/10234/24305>.

41. CHINER, E., [sin fecha]. Tema 6-Validez.pdf. [en línea]. [Consulta: 1 octubre 2022]. Disponible en: <https://rua.ua.es/dspace/bitstream/10045/19380/25/Tema%206-Validez.pdf>.

ANEXOS

Anexo 1: Matriz de operacionalización de variables de estudio.

VARIABLE	Definición conceptual	Dimensión	Indicador	Nivel de medición
Análisis de sentimientos	“Conjunto de técnicas informáticas que se utilizan para extraer, clasificar, comprender y evaluar opiniones expresadas en fuentes publicadas en Internet, comentarios en portales y otro contenido generado por el usuario” (Cambria & Hussain, 2012).	Modelo.	Precisión	Ordinal
Reseñas en restaurantes de comidas peruanas	Las reseñas son opiniones sinceras de los usuarios de nuestro negocio. Son ellos quienes dan información valiosa a otros usuarios que buscan un restaurante. (SocialwiBox, 2020).	Calificación	Nivel	Nominal
		Sentimientos (Polaridad)	Positivos	Nominal
			Negativos	
		Emociones	Alegría	Nominal
			Asco	
			Anticipación	
			Confianza	
			Enfado	
Miedo				
Sorpresa				
Tristeza				

Anexo 2: Matriz de consistencia

Titulo	Pregunta	Objetivo General	Preguntas específicas	Objetivos específicos	Variables	Definición	Definición Operacional	Dimensiones	Indicadores	Escala
Uso del análisis de sentimientos de reseñas en restaurantes de comidas peruanas	¿En qué consiste el uso del análisis de sentimientos de reseñas en restaurantes de comidas peruanas?	Realizar el análisis de sentimientos de reseñas en restaurantes de comidas peruanas.	¿Cuál es el resultado del análisis de sentimientos en cuanto a la calificación en restaurantes de comidas peruanas?	Determinar el análisis de sentimientos en cuanto a calificación en restaurantes de comidas peruanas.	Análisis de sentimientos	Conjunto de técnicas informáticas que se utilizan para extraer, clasificar, comprender y evaluar opiniones expresadas en fuentes publicadas en Internet, comentarios en portales y otro contenido generado por el usuario.	Es una variable que será construida a fin de mejorar la experiencia de los clientes. Las dimensiones de la presente variable son: modelo y opiniones.	Modelo.	Precisión	Ordinal
			¿Cuál es el resultado del análisis de sentimientos en cuanto a los sentimientos (Polaridad) en restaurantes de comidas peruanas?	Determinar el análisis de sentimientos en cuanto a sentimientos (Polaridad) en restaurantes de comidas peruanas.				Reseñas en restaurantes de comidas peruanas	Las reseñas son opiniones sinceras de los usuarios de nuestro negocio. Son ellos quienes dan información valiosa a otros usuarios que buscan un restaurante.	Es una variable que será evaluada con el fin de determinar el nivel de cada indicador. Las dimensiones de la presente variable son: ambiente, producto, atención y precios.
¿Cuál es el resultado del análisis de sentimientos en cuanto a las emociones en restaurantes de comidas peruanas?	Determinar el análisis de sentimientos en cuanto a las emociones en restaurantes de comidas peruanas.	Sentimientos (Polaridad)	Positivos							
			Emociones	Negativos						
		Alegría								
		Asco								
		Anticipación								
		Confianza								
		Enfado								
		Miedo								
		Sorpresa								
		Tristeza								

Anexo 3: Fichas de registro

Ficha de registro					
Investigador		Yamunaque Villegas Henry David		Tipo de prueba	DESCRIPTIVO
Institución		Universidad César Vallejo			
Dimensiones de estudio para reseñas en restaurantes de comidas peruanas		<ul style="list-style-type: none"> Calificación 			
Fecha de Inicio			Fecha final		
Variable		Indicador		Medida	Fórmula
Reseñas en restaurantes de comidas peruanas		<ul style="list-style-type: none"> Nivel 		Nominal	(Solo si hay un indicador con formulación)
#	N° Reseñas	Fuente	Ítem	Resultado	
1	287	Google Maps	Nivel 1	<ul style="list-style-type: none"> Lugar Pésima Pedido Ensalada Carne 	
			Nivel 2	<ul style="list-style-type: none"> Expectativa Carne Ensalada Cantidad Cobrar 	
			Nivel 3	<ul style="list-style-type: none"> Ceviche Lugar Platos Ambiente Agradable 	
			Nivel 4	<ul style="list-style-type: none"> Buena Lugar Piura Platos Ambiente 	
			Nivel 5	<ul style="list-style-type: none"> Bueno/Buena Lugar Carta Agradable Ambiente 	

Ficha de registro				
Investigador	Yamunaque Villegas Henry David		Tipo de prueba	DESCRIPTIVO
Institución	Universidad César Vallejo			
Dimensiones de estudio para reseñas en restaurantes de comidas peruanas	<ul style="list-style-type: none"> Sentimientos (Polaridad) 			
Fecha de Inicio		Fecha final		
Variable	Indicador	Medida	Fórmula	
Reseñas en restaurantes de comidas peruanas	<ul style="list-style-type: none"> Positivos Negativos 	Nominal	(Solo si hay un indicador con formulación)	
#	N° Reseñas	Fuente	Ítem	Resultado (Frecuencia)
1	287	Google Maps	Positivos	69.51
			Negativos	4.99

Ficha de registro					
Investigador		Yamunaque Villegas Henry David		Tipo de prueba	DESCRIPTIVO
Institución		Universidad César Vallejo			
Dimensiones de estudio para reseñas en restaurantes de comidas peruanas		<ul style="list-style-type: none"> Emociones 			
Fecha de Inicio				Fecha final	
Variable		Indicador		Medida	Fórmula
Reseñas en restaurantes de comidas peruanas		<ul style="list-style-type: none"> Alegría Asco Anticipación Confianza Enfado Miedo Sorpresa Tristeza 		Nominal	(Solo si hay un indicador con formulación)
#	N° Reseñas	Fuente	Ítem		Resultado (Frecuencia)
1	287	Google Maps	Alegría		25.27
			Asco		1.7
			Anticipación		6.6
			Confianza		28.74
			Enfado		0.68
			Miedo		1.28
			Sorpresa		2.97
			Tristeza		2.2

Anexo 4: Validación Experto – N°1

TABLA DE EVALUACIÓN DE EXPERTOS

- TABLA DE VALIDACIÓN DEL INSTRUMENTO DE EXPERTOS: Nivel, Polaridad Positiva, Polaridad Negativa, Alegría, Asco, Anticipación, Confianza, Enfadado, Miedo, Sorpresa y Tristeza.

I. DATOS GENERALES

Apellidos y Nombres del Experto: **Correa Calle, Teófilo Roberto**
 Título y/o Grado Académico: Maestro en Ingeniería de Sistemas con Mención en Tecnología de la Información y Comunicación e Ingeniero de Sistemas

Doctor () Magister (X) Ingeniero () Licenciado () Otro ().....

Universidad que labora: Universidad César Vallejo
 Fecha : 10/10/2022

TESIS : Uso del análisis de sentimientos de reseñas en restaurantes de comidas peruanas

Autor: Yamunaque Villegas, Henry David

Deficiente (0-20%) Regular (21-50%) Bueno (51-70%) Muy Bueno (71-80%) Excelente (81-100%)

Mediante la evaluación de expertos usted tiene la facultad de calificar la tabla de validación del instrumento involucrado mediante una serie de indicadores con puntuaciones especificadas en la tabla, con la valoración de 0% - 100%. Asimismo, se exhorta a las sugerencias de cambio de ítems que crea pertinente, con la finalidad de mejorar la coherencia de los indicadores para su valoración.

II. ASPECTOS DE VALIDACIÓN

INDICADOR	CRITERIO	VALORACION				
		0-20%	21-50%	51-70%	71-80%	81-100%
CLARIDAD	Es formulado con lenguaje apropiado.					90%
OBJETIVIDAD	Esta expresado en conducta observable.					95%
ACTUALIDAD	Es adecuado el avance, la ciencia y tecnología.					95%
ORGANIZACION	Existe una organización lógica.					90%
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.					95%
INTENCIONALIDAD	Adecuado para valorar los aspectos del sistema metodológico y científico.					95%
CONSISTENCIA	Está basado en aspectos teóricos y científicos.					95%
COHERENCIA	En los datos respecto al indicador.					95%
METODOLOGÍA	Responde al propósito de investigación.					95%
PERTENENCIA	El instrumento es adecuado al tipo de investigación.					95%
TOTAL						94%

III. PROMEDIO DE VALIDACIÓN

--

IV. OPCIÓN DE APLICABILIDAD

- (X) El instrumento puede ser aplicado, tal como está elaborado
 () El instrumento debe ser mejorado antes de ser aplicado

FIRMA DEL EXPERTO



TEOFILO ROBERTO
CORREA CALLE

Anexo 5: Validación Experto – N°2

TABLA DE EVALUACIÓN DE EXPERTOS

- TABLA DE VALIDACIÓN DEL INSTRUMENTO DE EXPERTOS: Nivel, Polaridad Positiva, Polaridad Negativa, Alegría, Asco, Anticipación, Confianza, Enfadado, Miedo, Sorpresa y Tristeza.

I. DATOS GENERALES

Apellidos y Nombres del Experto: **Jaramillo Atoche, Javier Eduardo**
 Título y/o Grado Académico: Maestro en Ingeniería de Sistemas con Mención en Tecnología de la Información y Comunicación e Ingeniero de Sistemas

Doctor () Magister (X) Ingeniero () Licenciado () Otro ().....

Universidad que labora: Universidad César Vallejo
 Fecha : 11/10/2022

TESIS : Uso del análisis de sentimientos de reseñas en restaurantes de comidas peruanas

Autor: Yamunaque Villegas, Henry David

Deficiente (0-20%) Regular (21-50%) Bueno (51-70%) Muy Bueno (71-80%) Excelente (81-100%)

Mediante la evaluación de expertos usted tiene la facultad de calificar la tabla de validación del instrumento involucrado mediante una serie de indicadores con puntuaciones especificadas en la tabla, con la valoración de 0% - 100%. Asimismo, se exhorta a las sugerencias de cambio de ítems que crea pertinente, con la finalidad de mejorar la coherencia de los indicadores para su valoración.

II. ASPECTOS DE VALIDACIÓN

INDICADOR	CRITERIO	VALORACIÓN				
		0-20%	21-50%	51-70%	71-80%	81-100%
CLARIDAD	Es formulado con lenguaje apropiado.					98
OBJETIVIDAD	Esta expresado en conducta observable.					100
ACTUALIDAD	Es adecuado el avance, la ciencia y tecnología.					100
ORGANIZACIÓN	Existe una organización lógica.					95
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.				80	
INTENCIONALIDAD	Adecuado para valorar los aspectos del sistema metodológico y científico.					100
CONSISTENCIA	Está basado en aspectos teóricos y científicos.					100
COHERENCIA	En los datos respecto al indicador.				80	
METODOLOGÍA	Responde al propósito de investigación.					100
PERTENENCIA	El instrumento es adecuado al tipo de investigación.					100
TOTAL						

III. PROMEDIO DE VALIDACIÓN

95.3

IV. OPCIÓN DE APLICABILIDAD

- (X) El instrumento puede ser aplicado, tal como está elaborado
 () El instrumento debe ser mejorado antes de ser aplicado

FIRMA DEL EXPERTO

Anexo 6: Validación Experto – N°3

TABLA DE EVALUACIÓN DE EXPERTOS

TABLA DE VALIDACIÓN DEL INSTRUMENTO DE EXPERTOS: Comodidad, Comidas mas solicitadas, Tipos de comidas, Reclamos, Satisfacción y Precio Idóneo.

I. DATOS GENERALES

Apellidos y Nombres del Experto: Altuna Tocco, Gerardo Arturo
 Título y/o Grado Académico: Maestro en Ingeniería de Sistemas con Mención en Tecnología de la Información y Comunicación e Ingeniero de Sistemas

Doctor () Magister (X) Ingeniero () Licenciado () Otro ().....

Universidad que labora: Universidad César Vallejo
 Fecha : 10/09/2022

TESIS : Propuesta de análisis de sentimientos de reseñas en restaurantes de comidas peruanas

Autor: Yamunaque Villegas, Henry David

Deficiente (0-20%) Regular (21-50%) Bueno (51-70%) Muy Bueno (71-80%) Excelente (81-100%)

Mediante la evaluación de expertos usted tiene la facultad de calificar la tabla de validación del instrumento involucrado mediante una serie de indicadores con puntuaciones especificadas en la tabla, con la valoración de 0% - 100%. Asimismo, se exhorta a las sugerencias de cambio de ítems que crea pertinente, con la finalidad de mejorar la coherencia de los indicadores para su valoración.

II. ASPECTOS DE VALIDACIÓN

INDICADOR	CRITERIO	VALORACION				
		0-20%	21-50%	51-70%	71-80%	81-100%
CLARIDAD	Es formulado con lenguaje apropiado.				75%	
OBJETIVIDAD	Esta expresado en conducta observable.				70%	
ACTUALIDAD	Es adecuado el avance, la ciencia y tecnología.				75%	
ORGANIZACION	Existe una organización lógica.				70%	
SUFICIENCIA	Comprende los aspectos de cantidad y calidad.				70%	
INTENCIONALIDAD	Adecuado para valorar los aspectos del sistema metodológico y científico.				70%	
CONSISTENCIA	Está basado en aspectos teóricos y científicos.				75%	
COHERENCIA	En los datos respecto al indicador.				70%	
METODOLOGIA	Responde al propósito de investigación.				75%	
PERTENENCIA	El instrumento es adecuado al tipo de investigación.				70%	
TOTAL					72.5%	

III. PROMEDIO DE VALIDACIÓN

72.5%

IV. OPCIÓN DE APLICABILIDAD

- (X) El instrumento puede ser aplicado, tal como está elaborado
 () El instrumento debe ser mejorado antes de ser aplicado


 GERARDO ARTURO ALTUNA TOCCO
 INGENIERO DE SISTEMAS
 Reg. CIP N° 204501
FIRMA DEL EXPERTO

Anexo 7: Validación

N°	Experto	Grado académico	Puntaje
1	Correa Calle, Teófilo Roberto	Magister	94
2	Jaramillo Atoche, Javier Eduardo	Magister	95
3	Altuna Tocto, Gerardo Arturo	Magister	72.5

Anexo 8: Test Retest

PRUEBA DE INSTRUMENTO	Emociones								Sentimientos		Score - N° Items	
	1	2	3	4	5	6	7	8	9	10	11	Detalle
P1	0,14	1,29	0,24	0,18	5,6	1,74	0,84	5,57	2,48	13,97	10	N° Reviews: 93 - Julio
P2	0,22	2,46	0,34	0,27	5,34	0,44	0,8	5,86	1,68	15,46	10	N° Reviews: 93 - Agosto
PEARSON =	0,98985844	99,38%										

Anexo 9: Prueba de normalidad – Dimensión Calificación

Resumen de procesamiento de casos

	Casos					
	Válido		Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
nivel	774	100,0%	0	0,0%	774	100,0%

Descriptivos

		Estadístico	Error estándar
nivel	Media	4,24	,039
	95% de intervalo de confianza para la media	Límite inferior	4,17
		Límite superior	4,32
	Media recortada al 5%	4,38	
	Mediana	5,00	
	Varianza	1,180	
	Desviación estándar	1,086	
	Mínimo	1	
	Máximo	5	
	Rango	4	
	Rango intercuartil	1	
	Asimetría	-1,537	,088
	Curtosis	1,680	,176

Pruebas de normalidad

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
nivel	,320	774	,000	,713	774	,000

a. Corrección de significación de Lilliefors

Anexo 10: Prueba de normalidad – Dimensión Emociones

Resumen de procesamiento de casos

	Válido		Casos Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
emociones	774	100,0%	0	0,0%	774	100,0%

Descriptivos

		Estadístico	Error estándar
emociones	Media	,6275	,01631
	95% de intervalo de confianza para la media	Límite inferior	,5955
		Límite superior	,6595
	Media recortada al 5%	,6328	
	Mediana	,8284	
	Varianza	,206	
	Desviación estándar	,45381	
	Mínimo	,00	
	Máximo	1,50	
	Rango	1,50	
	Rango intercuartil	1,00	
	Asimetría	-,516	,088
	Curtosis	-1,394	,176

Pruebas de normalidad

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
emociones	,245	774	,000	,743	774	,000

a. Corrección de significación de Lilliefors

Anexo 11: Prueba de normalidad – Dimensión Sentimientos (Polaridad)

Resumen de procesamiento de casos

	Válido		Perdidos		Total	
	N	Porcentaje	N	Porcentaje	N	Porcentaje
sentimientos	774	100,0%	0	0,0%	774	100,0%

Descriptivos

		Estadístico	Error estándar
sentimientos	Media	,7222	,01611
	95% de intervalo de confianza para la media	Límite inferior	,6906
		Límite superior	,7538
	Media recortada al 5%	,7469	
	Mediana	1,0000	
	Varianza	,201	
	Desviación estándar	,44819	
	Mínimo	,00	
	Máximo	1,00	
	Rango	1,00	
	Rango intercuartil	1,00	
	Asimetría	-,994	,088
	Curtosis	-1,014	,176

Pruebas de normalidad

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
sentimientos	,455	774	,000	,560	774	,000

a. Corrección de significación de Lilliefors

Anexo 12: Desarrollo de MVP

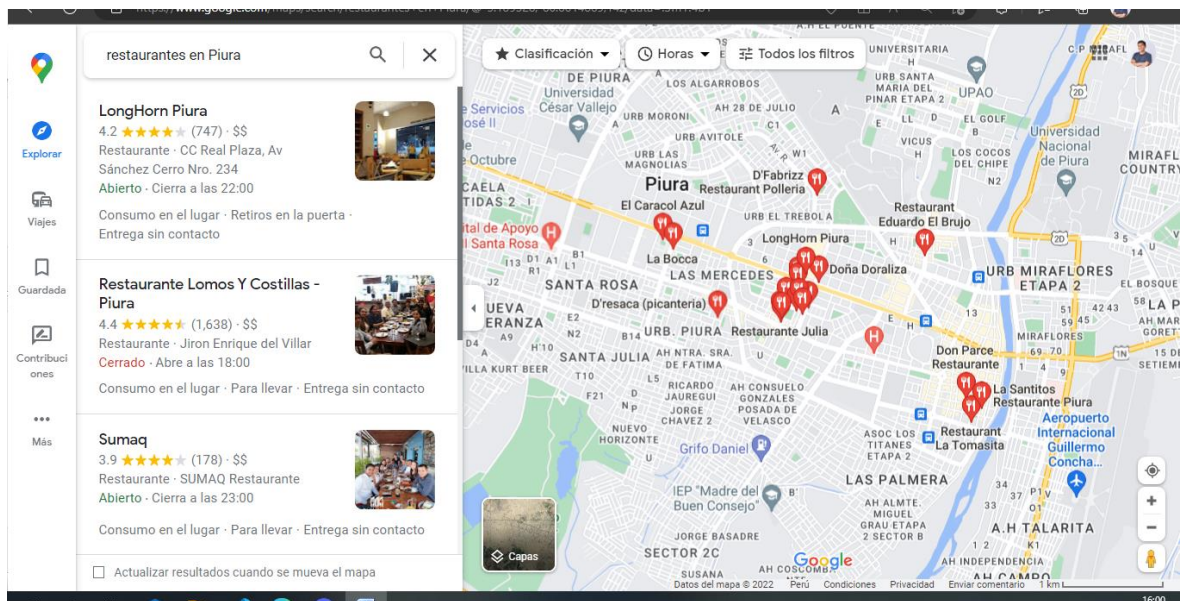
En el desarrollo de este prototipo en el cual se logró realizar el uso del Análisis de sentimientos en reseñas de restaurantes de comida peruana, la metodología aplicada para llevar a cabo este análisis de sentimientos se resume en la siguiente figura:



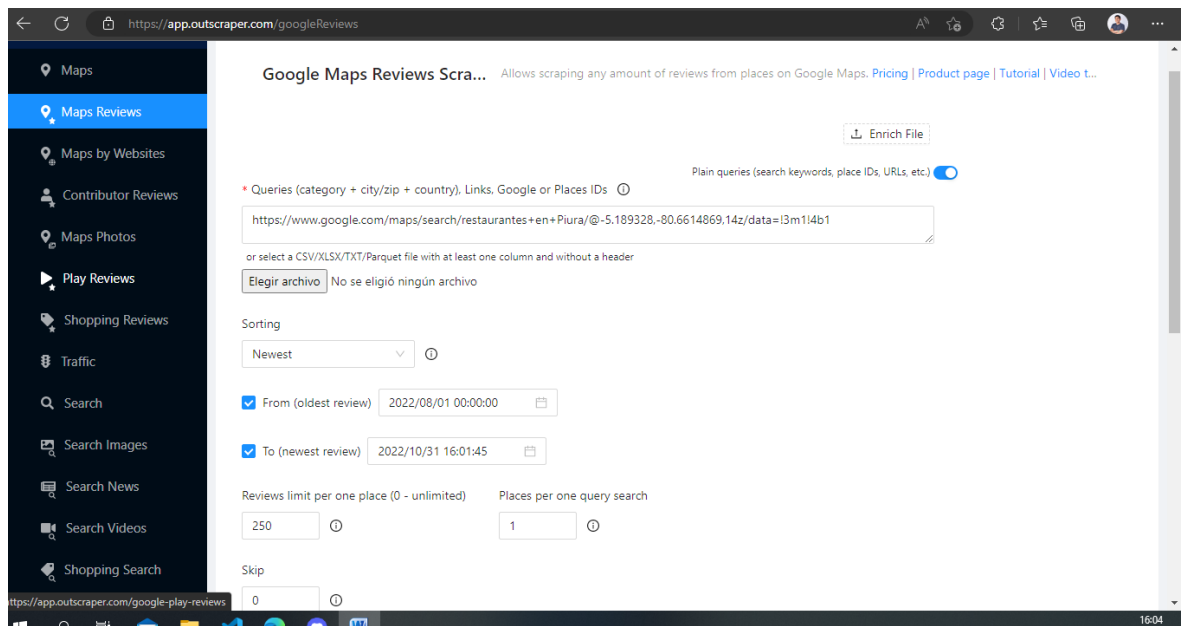
A continuación, se explica en detalle cuando de las etapas desarrolladas:

Recolección de reseñas: Se escogieron reseñas de restaurantes de comida peruana, para obtener tal data, se usó un programa apto para web scraping en el cual se pudo recopilar la data de los últimos 3 meses.

- Primeramente se obtuvo el link de la página (Google Maps) don las respectivas coordenadas.



- Se usó la herramienta de web scraping “Outscrapet”, el cual seleccionamos extraer reviews o reseñas de google maps.



Pre Procesamiento: Los datos recopilados se sometieron a una serie de técnicas que permitieron limpiar y reducir características de los textos, así facilitando el análisis de sentimientos. Las técnicas aplicadas en este caso fueron: filtrado de textos, tokenización, transform cases y eliminación de stopwords.

- Instalación de paquetes a usar:

```
install.packages("tidyverse")
library(tidyverse)

install.packages("data.table")
library(data.table)

install.packages("base")
library(base)

install.packages("stringi")
library(stringi)
```

- Cargar datos a pre procesar.

```
# obtener review a pre procesar
reviews<-fread('C:\\Users\\Henry\\Documents\\R\\mining\\text_mining\\reviews.csv', encoding='UTF-8')
```

- Función para el proceso


```

# Funcion para limpiar y tokenizar
limpiar_tokenizar <- function(texto){
  # El orden de la limpieza no es arbitrario
  Encoding(texto) <- "UTF-8"
  # Se convierte todo el texto a minusculas
  nuevo_texto <- tolower(stri_trans_general(texto,"Latin-ASCII"))
  # Eliminación de paginas web (palabras que empiezan por "http." seguidas de cualquier cosa
  # que no sea un espacio)
  nuevo_texto <- str_replace_all(nuevo_texto,"http\\s*", "")
  # Eliminación de signos de puntuación
  nuevo_texto <- str_replace_all(nuevo_texto,"[[:punct:]]", " ")
  # Eliminación de numeros
  nuevo_texto <- str_replace_all(nuevo_texto,"[[:digit:]]", " ")
  # Eliminación de emoticones
  nuevo_texto <- iconv(nuevo_texto, "UTF-8", "UTF-8",sub='') #iconv(nuevo_texto, to='ASCII//TRANSLIT'
  # Eliminación de espacios en blanco - multiples
  nuevo_texto <- str_replace_all(nuevo_texto,"[\\s]+", " ")
  # Tokenización por palabras individuales
  nuevo_texto <- str_split(nuevo_texto, " ")[[1]]
  # Eliminación de tokens con una longitud < 4
  nuevo_texto <- keep(.x = nuevo_texto, .p = function(x){str_length(x) > 4})
  # Convertir vector en string
  nuevo_texto <- paste(nuevo_texto, collapse = " ")
  return(nuevo_texto)
}

```

- Ejecutar función

```

# Limpiar_tokenizar
reviews_r <- reviews %>% mutate(pre_review = map(.x = review,
                                                .f = limpiar_tokenizar))

# Seleccionar columnas a usar
reviews_s <- reviews_r %>% select(id_review, pre_review)
# Exportar excel
fwrite(reviews_s, 'preprocessed_text.csv', row.names=FALSE)

```

Análisis de sentimientos: Después de haber procesado previamente los datos se realizó el análisis de sentimientos utilizando un clasificador, el cual etiquetara el sentimiento expresado en cada reseña.

- Se obtienen los datos a procesar

```

install.packages("data.table")
library(data.table)

# GET DATA
#GENERAL
reviews_df<-fread('C:\\Users\\Henry\\Documents\\reviews_r\\data\\data\\reviews.csv',encoding='UTF-8')
reviews_if<-fread('C:\\Users\\Henry\\Documents\\reviews_r\\preprocessed_text.csv',encoding='UTF-8')
reviews_df<-merge(reviews_df, reviews_if, by='id_review')
restaurants_df<-fread('C:\\Users\\Henry\\Documents\\reviews_r\\data\\data\\restaurants.csv', encoding='UTF-8')
nrc<-fread('C:\\Users\\Henry\\Documents\\reviews_r\\data\\data\\spanish-nrc\\lexico_nrc.csv', encoding = "UTF-8")

```

- Se instalan paquetes a usar

```

install.packages("syuzhet")
install.packages("lubridate")
install.packages("ggplot2")
install.packages("scales")
install.packages("reshape2")
install.packages("dplyr")
install.packages("qdap")
library(syuzhet)
library(lubridate)
library(ggplot2)
library(scales)
library(reshape2)
library(dplyr)
library(qdap)

install.packages("snowballc")
install.packages("tm")
install.packages("syuzhet")

library(snowballc)
library(tm)
library(syuzhet)

```

- Se declaró las emociones que se van tratar, como también cual es el sentimiento asociado (Polaridad) y la función a usar.

```

emotion<-c('anger','anticipation','disgust','fear','joy','sadness','surprise','trust')
sentiment<-c('negative','positive')
get_probs_nrclexicon<-function(x){
  x<-get_nrc_sentiment(char_v = x, language = "spanish")
  p_em<-as.numeric(x[,emotion])/sum(as.numeric(x[,emotion]))
  p_ss<-as.numeric(x[,sentiment])/sum(as.numeric(x[,sentiment]))
  p_em<-ifelse(is.na(p_em),rep(0,8),p_em)
  p_ss<-ifelse(is.na(p_ss),rep(0,2),p_ss)
  return(c(p_em,p_ss))
}

reviews_df$prob_nrc<-lapply(reviews_df$pre_review,get_probs_nrclexicon)
j<-0
for (em in c(emotion,sentiment)){
  j<-j+1
  reviews_df[,em]<-sapply(reviews_df$prob_nrc, function(x) x[j])
}

```

Clustering: Después de haber procesado los datos, se realizó el proceso de agrupar y separar según lo obtenido por el análisis de sentimientos, se utilizó el algoritmo de K.Means.

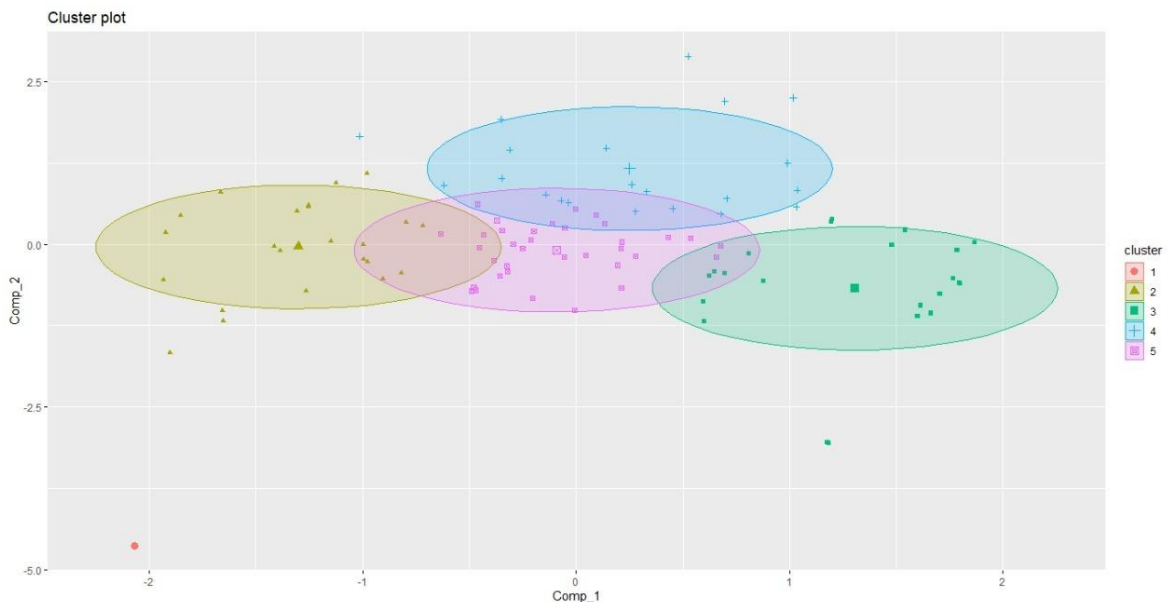
- Función para el clustering

```

get_clustering<-function(scores,s){
  scores<-scale(scores)
  SC<-data.table()
  for (k in 2:12){
    km<-kmeans(scores,k)
    SC<-rbind(SC,cbind(k,km$tot.withinss,km$betweenss,km$totss))
  }
  colnames(SC) <- c("n","SCI","SCE","SCT")
  print(SC)
  kmeans_fc<-function(x,k) kmeans(x, k, iter.max=50)
  fviz_nbclust(scores,kmeans_fc, method = "wss",linecolor = "red")>%print()
  fviz_nbclust(scores,kmeans_fc, method = 'silhouette',linecolor = "blue")>%print()
  fviz_nbclust(scores,kmeans_fc, method = "gap_stat",linecolor = "black")>%print()
  ResOp <- kmeans(scores,s)
  cluster <- ResOp$cluster %>% as.data.table()
  colnames(cluster) <- c("cluster")
  fviz_cluster(ResOp, data = scores,geom = 'point',repel = TRUE,
               palette='rainbowcols',ellipse.type = "euclid")>%print()
  return(cbind(scores,cluster) %>% as.data.table())
}

df<-get_clustering(pca_df,5)

```



Visualización: Luego de haber obtenido los resultados, se mostraran de manera intuitiva, ya sea en gráficos generados por R o gráficos en dashboard de Google Data Studio.

- Se exportan data frames de r studio a Excel, para usarlos en google data studio como fuente de datos.

```

#-----
# save this results
df<-cbind(all_df,df)
fwrite(df,'restaurants_new_a.csv',row.names=FALSE)
fwrite(reviews_df,'reviews_new_a.csv',row.names=FALSE)
#-----

```

- Función para el gráfico de nubes con las palabras más usadas en las reviews.

```
install.packages("tidytext")
library(tidytext)
install.packages("wordcloud2")
library(wordcloud2)

reviews_df<-merge(reviews_df,df%>%select(id,name,district,cluster),by.x='service',by.y='id')
Stopwords <- get_stopwords(language = 'spanish')%>%
  add_row(get_stopwords(language = 'english'))%>%
  add_row(word=c('pre_review','Length','dtype','q','object','local',
                'rapido','comida','restaurante','comer','rico',
                'servicio','plato','recome','recomendar','recomendarr',
                'recomendarndar','atencion','atender','excelente',
                'sabor','delicia','calidad','ano','and','precio','x')) #particular stopwords
words_df<-reviews_df%>%select(pre_review,id_review,score,service,district,cluster)%>%
  unnest_tokens(word, pre_review)%>%anti_join(Stopwords, by = "word")

plot_wordcloud<-function(x){
  x%>%count(word, sort = TRUE)%>%filter(n>10)%>%
  wordcloud2(size=1.6, minSize = 0.9, color='random-light', backgroundColor="black",
             shape="diamond", fontFamily="Comic Sans MS")}
plot_wordcloud(words_df)
```



- Método para el gráfico de las palabras más frecuentes según emociones, sentimiento (polaridad), clúster y nivel.

```

ngram_vars<-function(tokens,nrc,txt,c1) {
  if (txt%in%c('emotion','sentiment')) {tokens<-tokens %>%inner_join(nrc, "word")}
  if (txt=='cluster') tokens<-tokens%>%na.omit()
  tokens%>%count(get(txt), word, sort=T)%>% # The variable have name "get(txt)" now in each case
  group_by(get('get(txt)')) %>%
  arrange(desc(n))%>% slice(1:10) %>%
  ggplot(aes(x=reorder(word, n), y=n)) +
  geom_col(fill=c1,color='black', show.legend = F) +
  facet_wrap(~get('get(txt)'), scales = "free_y") +
  coord_flip() +
  theme_dark() +theme(plot.title = element_text(face = "bold"))+
  labs(x="", y="", title=paste("Palabras mas frecuentes - ",txt))
}
nrc$emotion<-nrc$sentiment
get_ngrams<-function(tokens){
  p1<-ngram_vars(tokens,
    nrc%>%filter(sentiment %in% c('negativo','positivo')),
    'sentiment','green')
  p2<-ngram_vars(tokens,
    nrc%>%filter(emotion %in% c('alegría','anticipación','asco',
    'confianza','enfado','miedo',
    'sorpresa','tristeza' )),
    'emotion','green')
  p3<-ngram_vars(tokens,'ok','score','green')
  p4<-ngram_vars(tokens,'ok','cluster','green')
  ggarrange(
    ggarrange(p1,p2,nrow=2,heights = c(1,2)),
    ggarrange(p3,p4,nrow=2),ncol=2)
}
get_ngrams(words_df)

```

- Grafico para la frecuencia de los sentimientos (polaridad) y emociones en cada review.

```

emocion <- df %>% select(anger, anticipation, disgust, fear, joy, sadness, surprise, trust, negative, positive)

#Creamos un data frame en el cual las filas serán las emociones
#y las columnas los puntajes totales

#Empezamos transponiendo emocion.df
emocion.df3 <- data.frame(t(emocion))

#Sumamos los puntajes de cada uno de las reviews para cada emocion
emocion.df3 <- data.frame(round(rowsums(emocion.df3),2))

#Nombramos la columna de puntajes como cuenta
names(emocion.df3)[1] <- "cuenta"

#Dado que las emociones son los nombres de las filas y no una variable
#transformamos el data frame para incluirlas dentro
emocion.df3 <- cbind("sentimiento" = rownames(emocion.df3), emocion.df3)

#Quitamos el nombre de las filas
rownames(emocion.df3) <- NULL

#Verificamos el data frame
print(emocion.df3)

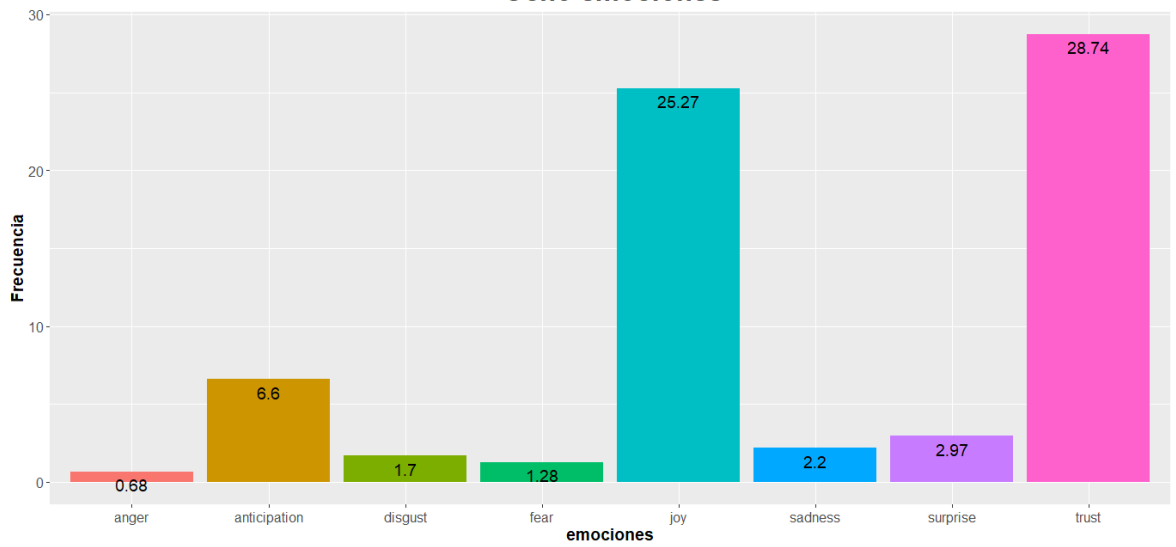
```

```

#Primer gráfico: se detallaran las 8 emociones con sus puntajes respectivos
sentimientos1 <- ggplot(emocion.df3[1:8,],
  aes(x = sentimiento,
      y = cuenta, fill = sentimiento)) +
  geom_bar(stat = "identity") +
  labs(title = "Análisis de sentimientos \n Ocho emociones",
    x = "emociones", y = "Frecuencia") +
  geom_text(aes(label = cuenta),
    vjust = 1.5, color = "black",
    size = 5) +
  theme(plot.title = element_text(hjust = 0.5),
    axis.text = element_text(size=12),
    axis.title = element_text(size=14,face = "bold"),
    title = element_text(size=20,face = "bold"),
    legend.position = "none")
print(sentimientos1)

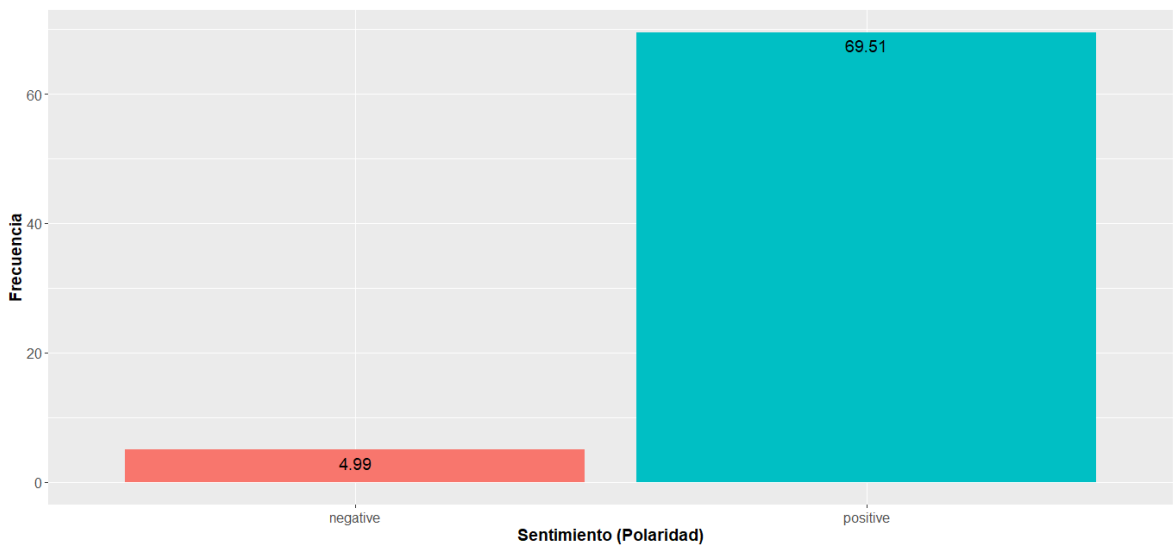
```

Análisis de sentimientos Ocho emociones



```
#Segundo gráfico: se detallan los puntajes de los sentimientos (Polaridad)
#positiva y negativa
sentimientos2 <- ggplot(emocion.df3[9:10,],
                        aes(x = sentimiento,
                            y = cuenta, fill = sentimiento)) +
  geom_bar(stat = "identity") +
  labs(title = "Análisis de sentimiento \n Polaridad",
        x = "Sentimiento (Polaridad)", y = "Frecuencia") +
  geom_text(aes(label = cuenta),
            vjust = 1.5, color = "black",
            size = 5) +
  theme(plot.title = element_text(hjust = 0.5),
        axis.text = element_text(size=12),
        axis.title = element_text(size=14,face = "bold"),
        title = element_text(size=20,face = "bold"),
        legend.position = "none")
print(sentimientos2)
```

Análisis de sentimiento Polaridad



- También se pueden visualizar más métricas en el link de google data studio: <https://datastudio.google.com/reporting/278a33b4-fcfa-4117-ba11-42c3f078a035>



UNIVERSIDAD CÉSAR VALLEJO

**FACULTAD DE INGENIERÍA Y ARQUITECTURA
ESCUELA PROFESIONAL DE INGENIERÍA DE SISTEMAS**

Declaratoria de Autenticidad del Asesor

Yo, ALTUNA TOCTO GERARDO ARTURO, docente de la FACULTAD DE INGENIERÍA Y ARQUITECTURA de la escuela profesional de INGENIERÍA DE SISTEMAS de la UNIVERSIDAD CÉSAR VALLEJO SAC - PIURA, asesor de Tesis titulada: "Uso del análisis de sentimientos de reseñas en restaurantes de comidas peruanas.", cuyo autor es YAMUNIQUE VILLEGAS HENRY DAVID, constato que la investigación tiene un índice de similitud de 19.00%, verificable en el reporte de originalidad del programa Turnitin, el cual ha sido realizado sin filtros, ni exclusiones.

He revisado dicho reporte y concluyo que cada una de las coincidencias detectadas no constituyen plagio. A mi leal saber y entender la Tesis cumple con todas las normas para el uso de citas y referencias establecidas por la Universidad César Vallejo.

En tal sentido, asumo la responsabilidad que corresponda ante cualquier falsedad, ocultamiento u omisión tanto de los documentos como de información aportada, por lo cual me someto a lo dispuesto en las normas académicas vigentes de la Universidad César Vallejo.

PIURA, 17 de Diciembre del 2022

Apellidos y Nombres del Asesor:	Firma
ALTUNA TOCTO GERARDO ARTURO DNI: 02715287 ORCID: 0000-0002-8311-4788	Firmado electrónicamente por: GALTUNATO el 17- 12-2022 23:06:40

Código documento Trilce: TRI - 0493816