



**FACULTAD DE POSTGRADO  
TRABAJO FINAL DE GRADUACIÓN**

**ANALISIS DE SENTIMIENTO BASADO EN LOS ASPECTOS  
DE LAS RESEÑAS DE UNA CADENA DE RESTAURANTES EN  
HONDURAS**

**SUSTENTADO POR:  
ROY EDUARDO COELLO VALLECILLO**

**PREVIO A INVESTIDURA DEL TÍTULO DE  
MÁSTER EN ANALÍTICA DE NEGOCIOS**

**SAN PEDRO SULA, CORTÉS, HONDURAS, C.A.**

**OCTUBRE 2023**

**UNIVERSIDAD TECNOLÓGICA CENTROAMERICANA  
UNITEC**

**FACULTAD DE POSTGRADO  
AUTORIDADES UNIVERSITARIAS**

**RECTORA  
ROSALPINA RODRÍGUEZ GUEVARA**

**PRORECTOR/ SECRETARIO GENERAL  
ROGER MARTÍNEZ MIRALDA**

**VICERRECTOR ACADÉMICO  
JAVIER ABRAHAM SALGADO LEZAMA**

**DIRECTORA NACIONAL DE POSTGRADO  
ANA DEL CARMEN RETALLY**

**ANALISIS DE SENTIMIENTO BASADO EN LOS ASPECTOS  
DE LAS RESEÑAS DE UNA CADENA DE RESTAURANTES EN  
HONDURAS**

**TRABAJO PRESENTADO EN CUMPLIMIENTO DE LOS  
REQUISITOS EXIGIDOS PARA OPTAR AL TÍTULO DE  
MÁSTER EN ANALÍTICA DE NEGOCIOS**

**ASESOR METODOLÓGICO  
ALBA GABRIELA GARAY ROMERO**

**ASESOR TEMÁTICO  
KEVIN EDUARDO FUNÉZ FUNÉZ**

**MIEMBROS DE LA TERNA:  
MARIO ALBERTO GALLO SANDOVAL  
HENRY DIOMEDES OSORTO NÚÑEZ  
ERICK LEONARDO IZAGUIRRE ARDON**

# DERECHOS DE AUTOR

© Copyright 2023

ROY EDUARDO COELLO VALLECILLO

Todos los derechos son reservados.

## FACULTAD DE POSTGRADO

# ANÁLISIS DE SENTIMIENTO BASADO EN LOS ASPECTOS DE LAS RESEÑAS DE UNA CADENA DE RESTAURANTES EN HONDURAS

**AUTOR:**

**ROY EDUARDO COELLO VALLECILLO**

### RESUMEN EJECUTIVO

El presente trabajo tiene como propósito realizar un análisis del sentimiento y clasificación de aspectos en reseñas obtenidas desde Google Maps de la cadena de restaurantes Denny's localizados en Honduras. El análisis del sentimiento es el proceso automático de identificar si la retroalimentación de un cliente es positiva o negativa, obteniendo una única fuente que permite evaluar las reseñas por contenido y contexto, no solo por la calificación general. El análisis de sentimiento basado en aspectos se puede utilizar para analizar los comentarios de los clientes asociándolos con diferentes aspectos del restaurante. En este trabajo, se implementaron diferentes modelos desarrollados en Python. Se obtuvo las reseñas haciendo uso de la plataforma Apify. Se realizó un preprocesamiento para obtener un conjunto de datos más legible para el algoritmo al momento de realizar la clasificación del sentimiento y la categorización de aspectos culinarios definidos como: precio, comida, lugar y servicio. El modelo con una mejor ponderación F1-Score fue el de regresión logística (LR), con un 95.56%. Los resultados muestran que la clasificación del sentimiento varía según los aspectos del restaurante. Se identificó que el aspecto del 'Lugar' es el mejor calificado con un 93% de positividad, seguidamente del 'Precio' (90%), 'Comida' (65%) y 'Servicio' (64%). Se recomienda a negocios, con una presencia activa en línea, a hacer el uso de métodos de análisis de sentimientos en las interacciones que obtienen en las redes sociales. Esta es una información valiosa que ayudara a identificar tendencias y tomar decisiones de negocio certeras que incrementen el nivel de lealtad en sus clientes y sus ganancias.

**Palabras clave:** análisis de sentimiento, aprendizaje automático, categorización de aspectos, LDA, regresión logística



## **GRADUATE SCHOOL**

# **SENTIMENT ANALYSIS BASED ON ASPECTS OF THE REVIEWS OF A RESTAURANT CHAIN IN HONDURAS**

**AUTHOR:**

**ROY EDUARDO COELLO VALLECILLO**

### **ABSTRACT**

The purpose of this work is to conduct a sentiment analysis and classification of aspects in reviews obtained from Google Maps of the Denny's restaurant chain located in Honduras. Sentiment analysis is the automatic process of identifying whether a customer's feedback is positive or negative, obtaining an only source that allows reviews to be evaluated by content and context, not just the overall rating. Aspect-based sentiment analysis can be used to analyze customer comments by associating them with distinct aspects of the restaurant. In this work, different models developed in Python were implemented. The reviews were obtained using the Apify platform. Preprocessing was conducted to obtain a more readable data set for the algorithm when performing sentiment classification and categorization of culinary aspects defined as: price, food, place, and service. The model with the best F1-Score was logistic regression (LR), with 95.56%. The results show that the classification of sentiment varies by the restaurant's aspects. It was identified that the aspect of 'Place' is the best rated with 93% positivity, followed by 'Price' (90%), 'Food' (65%) and 'Service' (64%). Any business, with an active online presence, is recommended to use sentiment analysis methods in the interactions they obtain on social networks. This is valuable information that will help identify trends and make accurate business decisions that increase the level of customer loyalty and profits.

**Key words:** aspect categorization, LDA, logistic regression, machine learning, sentiment analysis

## **DEDICATORIA**

Dedico este trabajo en primer lugar a Dios. El me dio fuerzas, sabiduría e inteligencia para seguir desarrollándome y continuar mis estudios.

Dedico este trabajo a mi familia, por el amor y apoyo constante que me han brindado.

Sin Dios y mi familia, no hubiese llegado tan lejos.

## **AGRADECIMIENTOS**

Agradezco primeramente a Dios, por estar siempre a mi lado. A mi familia, con la que se puedo contar siempre para recibir apoyo. Agradezco día a día su constante presencia en mi vida y el apoyo moral, económico o de cualquier índole que he recibido. Soy quien soy por cada uno de ustedes.

También agradezco a la Universidad Tecnológica Centroamérica (UNITEC) por brindarme su apoyo técnico, logístico en formación profesional y prepararme para ser mejores profesionales en el futuro.

# ÍNDICE DE CONTENIDO

<b>CAPITULO I. PLANTEAMIENTO DE LA INVESTIGACION .....</b>	<b>1</b>
1.1 INTRODUCCIÓN.....	1
1.2 ANTECEDENTES DEL PROBLEMA.....	2
1.3 DEFINICIÓN DEL PROBLEMA .....	3
1.3.1 PREGUNTAS DE INVESTIGACIÓN.....	3
1.4 OBJETIVOS.....	4
1.4.1 OBJETIVO GENERAL.....	4
1.4.2 OBJETIVOS ESPECIFICOS.....	4
1.5 JUSTIFICACIÓN .....	4
<b>CAPITULO II. MARCO TEORICO.....</b>	<b>6</b>
2.1 ANALISIS DE LA SITUACIÓN ACTUAL .....	6
2.1.1 ANALISIS DE DATOS Y LA TRANSFORMACION DEL MARKETING.....	6
2.1.2 BRECHA DIGITAL EN HONDURAS .....	7
2.2 CONCEPTUALIZACIÓN .....	9
2.2.1 SENTIMIENTOS.....	9
2.2.2 MINERIA DE REDES SOCIALES .....	9
2.2.3 LATENT DIRICHLET ALLOCATION (LDA).....	10
2.2.4 BAG OF WORDS.....	10
2.2.5 ASPECTOS DE UN RESTAURANTE.....	11
2.3 TEORÍAS DE SUSTENTO .....	11
2.3.1 BASES TEÓRICAS.....	11
2.3.2 METODOLOGÍAS DESARROLLADAS POR OTROS INVESTIGADORES O EXPERTOS.....	14
<b>CAPITULO III. METODOLOGÍA DE LA INVESTIGACIÓN .....</b>	<b>16</b>
3.1 CONGRUENCIA METODOLÓGICA.....	16
3.1.1 MATRIZ METODOLÓGICA.....	16
3.1.2 OPERACIONALIZACIÓN DE LAS VARIABLES .....	18
3.1.3 HIPÓTESIS.....	20
3.2 ENFOQUE Y MÉTODOS.....	20
3.3 DISEÑO DE LA INVESTIGACIÓN .....	21
3.3.1 POBLACIÓN .....	21
3.3.2 MUESTRA .....	21

3.3.3 TÉCNICAS DE MUESTREO .....	21
3.4 TÉCNICAS, INSTRUMENTOS Y PROCEDIMIENTOS APLICADOS.....	22
3.4.1 TÉCNICAS.....	22
3.4.2 PROCEDIMIENTOS.....	22
3.4.3 INSTRUMENTOS .....	28
3.5 FUENTES DE INFORMACIÓN .....	29
3.5.1 FUENTES PRIMARIAS.....	29
3.5.2 FUENTES SECUNDARIAS.....	29
<b>CAPITULO IV. RESULTADOS Y ANÁLISIS .....</b>	<b>30</b>
4.1 DESCRIPCIÓN DEL RESTAURANTE .....	30
4.2 RECOLECCIÓN DE DATOS .....	30
4.3 NUBES DE PALABRAS DE LOS ASPECTOS .....	31
4.3.1 ASPECTO DE SERVICIO .....	31
4.3.2 ASPECTO DE COMIDA.....	31
4.3.3 ASPECTO DE PRECIO.....	32
4.3.4 ASPECTO DE LUGAR .....	32
4.4 RESULTADOS Y ANÁLISIS DE LAS TÉCNICAS APLICADAS .....	33
4.3.1 REGRESIÓN LOGÍSTICA (LR).....	33
4.3.2 MÁQUINAS DE VECTORES DE SOPORTE (SVM).....	34
4.3.3 REGRESIÓN LOGÍSTICA CON AJUSTE DE PARAMETROS (LR-ADJ) .....	35
4.5 RESULTADOS .....	36
<b>CAPITULO V. CONCLUSIONES Y RECOMENDACIONES.....</b>	<b>37</b>
5.1 CONCLUSIONES .....	37
5.2 RECOMENDACIONES.....	38
<b>CAPITULO VI. APLICABILIDAD.....</b>	<b>39</b>
6.1 TÍTULO.....	39
6.2 JUSTIFICACIÓN .....	39
6.3 ALCANCE DE LA PROPUESTA.....	40
6.4 DESCRIPCIÓN Y DESARROLLO A DETALLE DE LA PROPUESTA.....	40
6.4.1 DESCRIPCIÓN CLARA DEL QUE Y COMO.....	40
6.4.2 DESARROLLO DE TODOS LOS ELEMENTOS NECESARIOS .....	41
6.4 MEDIDAS DE CONTROL .....	42

6.5 GUÍA DE IMPLEMENTACIÓN PARA ANALISIS DE SENTIMIENTOS BASADOS EN ASPECTOS EN RESTAURANTES.....	42
6.5.1 RECOLECCIÓN DE DATOS .....	42
6.5.2 PREPROCESAMIENTO DE DATOS .....	42
6.5.3 APLICAR LDA PARA IDENTIFICAR LOS ASPECTOS.....	43
6.5.4 ASIGNACIÓN DEL SENTIMIENTO .....	44
6.6 CRONOGRAMA DE IMPLEMENTACIÓN Y PRESUPUESTO.....	45
6.5.1 PRESUPUESTO .....	45
6.5.2 CRONOGRAMA.....	46
<b>REFERENCIAS.....</b>	<b>47</b>
<b>ANEXOS.....</b>	<b>50</b>
ANEXO 1: CONJUNTO DE DATOS .....	50
ANEXO 2: MATRIZ DE CONFUSIÓN .....	50
ANEXO 3: RESULTADOS DE METRICA.....	50

## ÍNDICE DE ILUSTRACIONES

Ilustración 1 Esquema metodológico de investigación.....	20
Ilustración 2 Marco de investigación.....	22
Ilustración 3 Ejemplo de reseña procesada.....	23
Ilustración 4 Código Python del LDA.....	25
Ilustración 5 Resultados del LDA con parámetros establecidos.....	25
Ilustración 6 Formula del Puntaje F1.....	28
Ilustración 7 Nube de palabras - Aspecto de Servicio.....	31
Ilustración 8 Nube de palabras - Aspecto de Comida.....	31
Ilustración 9 Nube de palabras - Aspecto de Precio.....	32
Ilustración 10 Ilustración 10 Nube de palabras - Aspecto de Lugar.....	32
Ilustración 11 Resultados matriz de confusión de LR.....	33
Ilustración 12 Resultados de medición de LR.....	33
Ilustración 13 Resultados matriz de confusión de SVM.....	34
Ilustración 14 Resultados de medición de SVM.....	34
Ilustración 15 Resultados matriz de confusión de LR-ADJ.....	35
Ilustración 16 Ilustración 8 Resultados de medición de LR-ADJ.....	35
Ilustración 17 Ejemplo de preprocesamiento de datos usando Spacy en Python.....	43
Ilustración 18 Ejemplo de LDA para la clasificación de aspectos en Python.....	44
Ilustración 19 Ejemplo de Regresión Logística para la clasificación del sentimiento.....	44

## ÍNDICE DE TABLAS

Tabla 1 Inversión necesaria para cerrar brecha digital.....	8
Tabla 2 Matriz de congruencia metodológica.....	17
Tabla 3 Operacionalización de las variables.....	19
Tabla 4 Análisis PoS de una oración .....	24
Tabla 5 Categorización del algoritmo LDA a Servicio.....	26
Tabla 6 Categorización del algoritmo LDA a Comida .....	26
Tabla 7 Categorización del algoritmo LDA a Precio.....	27
Tabla 8 Categorización del algoritmo LDA a Lugar .....	27
Tabla 9 Conjunto de datos obtenidos.....	30
Tabla 10 Resultados del modelo .....	36
Tabla 11 Ejemplo de reseña procesada y clasificada.....	36
Tabla 12 Costo de equipo tecnológico.....	45
Tabla 13 Costo de software.....	45
Tabla 14 Costo de mobiliario.....	45
Tabla 15 Costo del recurso humano.....	45
Tabla 16 Cronograma de actividades.....	46

# **CAPITULO I. PLANTEAMIENTO DE LA INVESTIGACION**

## **1.1 INTRODUCCIÓN**

En el ambiente altamente competitivo en el que nos encontramos es imperativo para los negocios como ser restaurantes, cafés, bares, etc. estar al tanto de la retroalimentación recibida de sus clientes. Es lo más común por parte de los negocios realizar una investigación cualitativa para comprender a sus clientes objetivo y poder así darse una idea para tomar decisiones respecto al producto que desean comercializar. Al indagar sobre esto, se genera un análisis de sentimientos superficial. Saber cómo se siente los clientes sobre el producto y/o servicio provisto es imprescindible para su éxito y comprender de una forma detallada y estadística los sentimientos de la audiencia provee una ventaja sobre los competidores.

A través del aprendizaje automático y el análisis de texto, los algoritmos pueden clasificar las declaraciones como positivas, negativas y neutrales. Este trabajo realizara un análisis de sentimiento basado en aspectos o temas predefinidos de las reseñas, definidas como documentos de ahora en adelante, de los comensales dejados en Google Maps para obtener mejores bases que ayuden a diseñar estrategias de negocios más completas. Cada documento será preprocesado para su respectiva tokenización, eliminación de stopwords o palabras de vacío, convertir en minúsculas, remover caracteres no alfabéticos y lematización. Consecuentemente, cada documento se categorizará en 4 aspectos o temas definidos: Comida, Servicio, Precio y Lugar.

El trabajo hace uso del método LDA (Latent Dirichlet Allocation), ya que a través de este se puede identificar los temas latentes de un documento.

Se pretende que el algoritmo de análisis de sentimiento utilizado ahorrara tiempo valioso a la empresa organizando y clasificando los mensajes de los clientes de múltiples canales de comunicación (redes sociales, etc.).

Se aplicará un algoritmo de análisis de sentimiento basado en aspectos realizado en Python en la cadena de restaurantes Denny's ubicados Honduras. Con dicho análisis, el negocio en cuestión podrá aplicar medidas correctivas de ser necesario a su servicio y/o productos para así hacerse de un modelo de negocio más competitivo en un mercado saturado en la ciudad actualmente.

## **1.2 ANTECEDENTES DEL PROBLEMA**

El uso de las redes sociales a incrementado considerablemente en la última década. La proliferación de estas ha llegado a tal grado que cerca del 93% de los negocios poseen por lo menos una cuenta en las redes como Facebook, Twitter, etc. (Thota, 2018). Desde que la pandemia de COVID-19 azoto el mundo, se reporta un uso mayor de las redes sociales de parte de los consumidores proporcionando más interacciones en estas plataformas. (Samet, 2020).

Este incremento al uso de las redes sociales se traduce a un aumento en la disposición de los consumidores a compartir experiencias de consumo en línea. Con la tecnología disponible para analizar estos datos, se puede ofrecer a los gerentes de marketing una oportunidad de recopilar inteligencia de mercado sin precedentes (Erevelles, Fukawa, & Swayne, 2016). Según (Baquedano, 2022), el 48% de las empresas venden por internet. En el 2018, el 16.2% de las compañías informaron que usaban información de las redes sociales para la toma de decisiones. Al 2020, cerca del 48.3% lo hacían.

Si bien el análisis de sentimiento aplicado en las redes sociales presenta una oportunidad atractiva y potencialmente lucrativa para muchos negocios al interpretar grandes cantidades de texto recolectado en un ambiente natural de manera automática, hay preocupaciones sobre la precisión y practicidad de las técnicas actuales empleadas. (Goncalves, Araujo, Benevenuto, & Cha, 2013)

Además, las aprensiones sobre el volumen de los datos en línea, las fuentes de datos fragmentadas, el sesgo de contenido y la explotación del usuario han expuesto la técnica del análisis del sentimiento al escrutinio crítico.

Un estudio de mercado, especialmente en el rubro de los restaurantes, es esencial para la toma de decisiones del negocio. Los consumidores son parte fundamental de la actividad del mercado y realizar un estudio ofrece los medios para analizar el comportamiento y percepciones. Lo anterior se realiza usualmente siguiendo un enfoque cuantitativo o cualitativo. El cuantitativo implica usualmente la distribución de cuestionarios a una muestra representante de personas y posteriormente, se interpreta los resultados estadísticamente. Con la parte cualitativa, el objetivo es más sobre conseguir una perspectiva más profunda sobre un tema a través de grupos focales o entrevistas personales.

### 1.3 DEFINICIÓN DEL PROBLEMA

El primer paso en la cimentación de una marca con el mercado consumidor es tener presencia en las redes sociales. De ahí el plan radica en crear y fomentar interacciones con los seguidores de las páginas del negocio. Subir imágenes sencillas como fotos de los productos y/o servicios que venden, fotos del local (si aplica), hasta fotos haciendo marketing de nuevos productos o promociones con el objetivo de generar más ventas. Las redes sociales son el medio por el cual el negocio se hace notar frente a nuevos y potenciales mercados.

Comprender, no simplemente leer, lo que los consumidores comentan y discuten en las redes sociales ayudara al negocio a comprender el ‘por qué’ detrás de las acciones del cliente. ¿Por qué prefieren a la competencia sobre nuestro negocio? ¿Por qué el índice de retención de nuestros clientes es tan bajo?

Usaremos los conocimientos obtenidos de esta investigación en un restaurante ubicado en San Pedro Sula, Honduras. Se espera que, con esta información, el restaurante podrá sacar provecho de su presencia en las redes sociales – recibiendo de manera automática *insights* de sus consumidores categorizados en: Local, Ambiente, Precios, Comida, Servicio. El negocio recibirá datos estadísticos que le demuestren que es lo que sus consumidores piensan realmente de su negocio y partiendo de ahí la gerencia podrá tomar decisiones que ayuden a mejorar. Se obtendrá información que antes se invierte recursos para recopilar de las varias cuentas sociales de un negocio a tan solo un clic de botón.

#### 1.3.1 PREGUNTAS DE INVESTIGACIÓN

La principal interrogante es, ¿Es posible realizar un análisis de sentimiento con la información de las reseñas en Google Maps de una cadena de restaurantes en Honduras?

¿Cuáles son los aspectos más mencionados en las reseñas del restaurante?

¿Cuál es la polaridad de los aspectos según las reseñas del restaurante?

¿Cuáles son las oportunidades de mejora identificadas en el negocio?

## **1.4 OBJETIVOS**

### **1.4.1 OBJETIVO GENERAL**

Realizar mediante machine learning un análisis del sentimiento y categorización de la retroalimentación brindada por clientes de una cadena de restaurantes de comida rápida en Google Maps.

### **1.4.2 OBJETIVOS ESPECIFICOS**

1. Categorizar cada reseña identificada en cuatro aspectos culinarios predefinidos.
2. Identificar mediante el algoritmo de machine learning la polaridad de los aspectos o temas definidos en las reseñas de los comensales.
3. Determinar las oportunidades de mejora en los aspectos según el análisis del sentimiento de los comentarios brindados por los clientes.

## **1.5 JUSTIFICACIÓN**

Las redes sociales llegaron a revolucionar el campo de juego en el mercado para los pequeños y medianos negocios. En esencia, son una herramienta de marketing gratuita que la mayoría del mundo usa. Es la base de datos de potenciales clientes más grande que existe, cambiando minuto a minuto. Aprender a controlar esta herramienta, y usarla al beneficio del negocio, es el desafío que muchos negocios están enfrentando y muchos fallan en lograrlo.

Un negocio puede tener más de una red social y tratar de recopilar, limpiar, ordenar y analizar cada comentario y mensaje recibido de la clientela es una tarea de proporciones épicas: requiere una extensa inversión de recursos y tiempo poder hacer un análisis completo de los comentarios recibidos y muchas veces dicho análisis es erróneo ya que en ciertos casos resulta imposible descifrar que es lo que un consumidor pretendía o sentía al momento de realizarlo y enviarlo.

Como se mencionó anteriormente, las redes sociales son una gran oportunidad para la recolección de opiniones que los clientes dejan. Sin embargo, es difícil hacer sentido de dicha información no estructurada. El cliente hace uso de un lenguaje natural para dejar la reseña u opinión lo que lo hace parecer como si se tuviese una entrevista de enfoque individual por cada

una. En casos así, potencialmente el negocio trataría con cientos o miles de personas dejando su opinión haciendo uso de su lenguaje natural – obteniendo así el muestreo necesario para un estudio cuantitativo, manteniendo el enfoque profundo de información específica a un estudio cualitativo.

Para obtener los resultados estadísticos que se obtendrían de un estudio cuantitativo es necesario encontrar una manera de estructurar la información. Aquí es donde el procesamiento de lenguaje natural (NLP, por sus siglas en inglés) y aprendizaje automático (ML, por sus siglas en inglés) entran en juego. El NLP ofrece la manera de estructurar estos datos y extraer características significativas de ellos, mientras que ML descubre los patrones detrás de estas características. Todo esto es hecho de manera automática, ocupando un valor monetario mucho menor que un estudio clásico. Una vez que los modelos estén desarrollados y habilitados, los resultados pueden extraerse en tiempo real. Esto hace que ya no haya ningún atraso entre un evento, el negocio observando dicho evento y ordenando una investigación y la información siendo recolectada, desarrollada, interpretada y visualizada.

## CAPITULO II. MARCO TEORICO

### 2.1 ANALISIS DE LA SITUACIÓN ACTUAL

El término ‘Análisis de sentimiento’, aparece por primera vez en un artículo publicado en el 2003 por los investigadores de IBM Nasukawa y Yi, llamado ‘Sentiment Analysis: capturing favorability using natural language processing’. Este campo de investigación también se denomina ‘minería de opinión’, un término que apareció por primera vez en el mismo año en el artículo ‘Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews’ (Liu, 2020).

Debido a que este tema se enfoca mayormente en textos, el análisis de sentimiento es considerado parte del PLN (procesamiento de lenguaje natural). Este campo del PLN ha comenzado a ganar popularidad en los últimos años debido al incremento en el uso de las redes sociales. Este incremento, lleva a una gran cantidad de información no estructurada en la forma de textos dejadas por los usuarios (opiniones, reseñas, blogs, etc.). Otro factor es debido a la continua mejora en los algoritmos de PLN y de recolección de información que han facilitado de gran manera estos análisis.

#### 2.1.1 ANALISIS DE DATOS Y LA TRANSFORMACION DEL MARKETING

El análisis del consumidor se encuentra en el cruce de *Big Data* y el comportamiento del consumidor. Mediante la información proveída sobre el comportamiento de los consumidores; los especialistas en marketing transforman esa información en una ventaja en el mercado. La analítica se refiere generalmente a las herramientas utilizadas para descubrir patrones en esa información. (Statista, 2021)

Gracias al aumento del uso del internet, los negocios generan más información de la que son capaces de utilizar o de la que saben cómo utilizar. La diferencia de hoy en día es el increíble volumen, variedad y velocidad de datos que se obtenidos de los consumidores resultando en una revolución de la Big Data – una revolución que conducen a formas completamente nuevas e innovadoras de comprender el comportamiento del consumidor lo que permite formular estrategias de marketing mucho más efectivas de lo que era antes posible. (BID, 2021) En esta investigación, el análisis de Big Data del consumidor se define como la extracción de información oculta sobre el comportamiento del consumidor y la explotación de esa información a través de la interpretación.

El uso del internet y la brecha digital en los países cada vez se hace más pequeña lo que ocasiona que la información obtenida del mismo sea cada vez mayor, más completa e inexplicable frente al limitado recurso humano, ocasionando dificultades para descifrarla e interpretarla en un mercado desconocido. Para poder sacar el mejor provecho de este nuevo capital, las empresas y negocios deben alocar una cantidad apropiada de recursos humanos y económicos. Con la big data se trae un cambio importante al método científico, de encajar datos a teorías preconcebidas del mercado al uso de datos para crear nuevas teorías.

### 2.1.2 BRECHA DIGITAL EN HONDURAS

Si bien la Big Data es considerada como el nuevo capital en el marketing digital – Honduras se considera como uno de los países con mayor brecha digital en la región según estudio realizado por la Universidad Autónoma de Honduras (Observatorio Demográfico Universitario, 2021). Según este estudio, referente a la cantidad de hogares conectados a internet hasta el 2019, Honduras se posicionaba en el penúltimo lugar con 19.3%, siendo Costa Rica y Panamá los países con mayor índice de hogares conectados a internet con 86.3% y 70.7% respectivamente. Según CONATEL, la mayoría de hondureños tiene acceso a internet mediante el uso de internet móvil, siendo la cantidad de usuarios más de 8 millones de hondureños, un aumento del 6.6% según datos del 2022 en comparación con el 2021. (Mesa Editorial, 2022).

Si bien Honduras y la región está atrasado con una alta brecha digital, la tendencia de la digitación y la importancia del análisis de datos es algo que no se puede frenar. El gobierno de Honduras está trabajando para tratar de disminuir esta brecha, brindando internet gratuito a más de 75 mil hogares con el objetivo de garantizar una educación a los niños que estudian desde casa. Todas estas acciones garantizan que el uso del internet llegue a lugares más remotos del país, ocasionando que un nuevo mercado y una nueva fuente de datos que antes era muy difícil obtener sea más accesible para las empresas que sepan aprovecharlo. (Secretaria Educacion, 2021).

“Los retrasos en la mejora de la conectividad y la digitalización en los países de América Latina y el Caribe han exacerbado drásticamente el impacto económico y social de la COVID-19. Pero esta realidad también ofrece una oportunidad histórica para reducir la desigualdad y generar empleos y un crecimiento económico sostenible”, dijo el presidente del BID, Mauricio Claver-Carone. Esta información se puede apreciar en la Tabla 1.

Tabla 1 Inversión necesaria para cerrar brecha digital

<b>Pais</b>	<b>Inversion necesaria para cerrar la brecha - movil</b>	<b>Inversion necesaria para cerrar la brecha fija</b>	<b>Estimacion de empleos directos creados</b>	<b>Porcentaje de inversion en areas rurales</b>
Brazil	\$4,057,805,055	\$17,758,097,269	3,727,628	31.00%
Mexico	\$3,790,612,056	\$8,013,328,140	3,048,264	42.90%
Colombia	\$2,006,378,307	\$3,803,526,628	1,519,161	41.30%
Argentina	\$1,406,288,593	\$3,285,952,468	1,045,961	20.70%
Peru	\$1,069,530,775	\$3,827,997,489	917,206	45.80%
Venezuela	\$1,129,186,788	\$2,974,694,670	905,420	34.60%
Guatemala	\$1,125,241,660	\$1,148,699,142	839,725	74.20%
Ecuador	\$666,634,444	\$1,233,195,390	520,673	62.80%
Haiti	\$595,543,939	\$378,651,741	476,061	70.20%
Honduras	\$501,252,454	\$852,678,253	392,007	68.50%

Fuente: (BID, 2021)

#### 2.1.2.1 STARLINK EN HONDURAS

Según (Diario La Prensa, 2023), en 2024 entrara a Honduras Starlink. Starlink es un servicio de la compañía SpaceX que provee internet satelital haciendo uso de constelaciones de satélites que sobrevuelan la tierra en baja orbita. Contando con más de 5000 satélites a lo que va de agosto 2023, hay planes de desplegar 12000 más y se espera expandirse hacia 42000 más adelante. A lo que va de septiembre 2023, el servicio cuenta con más 2 millones de subscriptores alrededor del mundo. (Starlink, n.d.).

Este servicio tiene presencia en los demás países centroamericanos y provee velocidades de descarga de alrededor de 100 megabits por segundo (mbps). Según la Comisión Nacional de Telecomunicaciones (CONATEL), la velocidad promedio para usuarios de internet fijo en Honduras es de 17.41 mbps y 21.88 mbps para usuarios móviles. Este servicio podrá llegar en partes de Honduras donde las empresas establecidas actualmente en el país no llegan y se espera que impulse de gran manera el cierre de la brecha digital en el país. Esto se espera que impulse en gran manera los mercados en línea del país, permitiendo que más personas participen y consigo traiga más percepciones de los sentimientos de los usuarios de productos y/o servicio.

## 2.2 CONCEPTUALIZACIÓN

### 2.2.1 SENTIMIENTOS

El propósito del análisis del sentimiento es identificar los sentimientos de un texto escrito. De este punto, (Scherer, 1984) propone una tipología diferente a los estados de afecto:

- Emoción: un episodio relativamente corto de respuestas sincrónicas a la evaluación de un evento interno o externo que se considera de mayor importancia (por ejemplo: ira, felicidad, vergüenza, tristeza, entusiasmo, etc.)
- Estado de ánimo: estado afectivo difuso de baja intensidad y larga duración (por ejemplo: alegre, depresivo, optimista, sombrío, apático. etc.)
- Postura interpersonal: una postura afectiva adoptada hacia otra persona en una interacción específica (por ejemplo: distante, fría, amistosa, solidaria, etc.)
- Actitud: creencia, preferencia y predisposición duraderas hacia objetos o personas (por ejemplo: me gusta, amo, odio, quiero, aprecio, etc.)
- Rasgo de personalidad: disposición de personalidad estable y tendencia de comportamiento típica (por ejemplo: ansioso, celoso, gruñón, imprudente, etc.)

En vista a esta tipología, el análisis de sentimiento detecta actitudes, ya que en el texto se trata con personas teniendo una cierta actitud (fuente) y un objetivo, una polaridad (positiva, negativa o neutral) y una intensidad de esa actitud.

Esta falta de un consenso total para definir el sentimiento es un indicador que los mismos son mixtos y multidimensionales. Son bastante dependientes del contexto en que se encuentran y complejos desde puntos de vistas sociales y lingüísticos. Identificarlos puede ser un reto para un humano, no se diga una computadora.

### 2.2.2 MINERÍA DE REDES SOCIALES

Definimos que las redes sociales rompen las barreras entre el mundo real y el mundo virtual. Podemos ahora integrar teorías sociales con métodos computacionales para estudiar como los individuos (átomos sociales) interactúan y como las comunidades (moléculas sociales) se forman. La singularidad de las redes sociales exige métodos novedosos de minería de datos que puedan efectivamente recolectar el contenido repleto de interacciones sociales. El estudio y el desarrollo

de estos métodos están bajo el dominio de la minería de redes sociales, una disciplina emergente bajo el cuidado de la minería de datos. (Moreno, Martelo, Corredor, Cifuentes, & Lopez, 2020)

La minería de redes sociales se define entonces como el proceso de representar, analizar y extraer patrones accionables de la información generada por los usuarios.

### 2.2.3 LATENT DIRICHLET ALLOCATION (LDA)

Asignación Latente de Dirichlet (ALD) o LDA por sus siglas en ingles es un modelo generativo no supervisado que categoriza las palabras que aparecen en documentos en grupos referidos como aspectos o temas. La idea básica del LDA es que un documento consiste en uno o varios aspectos o temas latentes donde cada uno consiste en una distribución de palabras que los identifica. El peso de cada tema varía entre documento a documento es algo relativo. (Mo, Kontonatsios, & Ananiadou, 2015)

### 2.2.4 BAG OF WORDS

El Bag of Words (BOW) es crear un vocabulario de los tokens (palabras) encontradas en todos los documentos y después representar cada documento como un vector que contiene cada token en el vocabulario y su frecuencia.

Usando TF-IDFS (Term Frequency-Inverse Document Frequency), se consigue el producto de dos estadísticas: la frecuencia de los términos (la cantidad de veces que un término aparece en un documento) y la inversa de la frecuencia del documento (que tan común es ese término a través de todos los documentos). Esto se calcula de la siguiente forma:

$$idf(t, D) = \log \frac{N}{|d \in D, t \in d|}$$

*Ecuación 1 TF-IDFS*

En donde N es la cantidad total de documentos y  $|d \in D, t \in d|$  denota el numero de documentos que el que termino aparece. (Mo, Kontonatsios, & Ananiadou, 2015)

Esto soluciona el problema de documentos más largos teniendo un promedio de frecuencias más grande que los documentos más cortos, aun cuando ambos hablan del mismo tema, y también disminuye el peso de palabras que se encuentran en muchos documentos y son menos informativas que aquellas que ocurren en una porción más pequeña del vocabulario.

## 2.2.5 ASPECTOS DE UN RESTAURANTE

(Ahmed, Asheq, Ahmed, Chowdhury, & Mostafa, 2022), indican que la calidad percibida del servicio y la equidad del precio tienen un efecto directo y significativo en la satisfacción del consumidor. Además, los resultados de la investigación revelan una asociación considerable entre la percepción de justicia de precios y el placer y la lealtad del consumidor. Además, los hallazgos indican que la calidad percibida del servicio y la equidad del precio tienen un efecto indirecto significativo en la lealtad del cliente a través del efecto mediador de la satisfacción del consumidor.

## 2.3 TEORÍAS DE SUSTENTO

### 2.3.1 BASES TEÓRICAS

El análisis de sentimiento es una técnica usada para analizar la retroalimentación de los clientes sobre marcas, productos y servicios que se encuentran en reseñas en línea o en plataforma de redes sociales. Existen diferentes enfoques para analizar el sentimiento de los textos, incluidos los enfoques basados en léxico y en aprendizaje automático. (Bezco, 2023)

En el enfoque basado en léxico se usa un diccionario de palabras para determinar el sentimiento de un texto. Cada palabra en el léxico se le asigna una puntuación que indica su sentimiento positivo o negativo. El sentimiento general del texto se determina sumando las puntuaciones de todas las palabras del texto. El beneficio de este enfoque es su simplicidad e interpretabilidad. Se puede acceder directamente al léxico, lo que significa que se puede inspeccionar, comprender, ampliarlo y modificarlo fácilmente. En comparación con el análisis de sentimiento basado en aprendizaje automático, el análisis de sentimientos basado en léxico es más interpretable porque es fácil identificar las palabras que afectan la puntuación final y actualizar la puntuación de cada palabra de ser necesario. Sin embargo, a medida que se utilizan nuevas palabras o frases, como jergas o acrónimos, para transmitir emociones, habría que reconfigurar constantemente el léxico para garantizar precisión. (Bezco, 2023)

En el enfoque de aprendizaje automático, se utilizan algoritmos para aprender las características del texto relevantes para el sentimiento. El algoritmo entrena con un conjunto de datos de oraciones positivas, neutrales y negativas, y luego se usa para clasificar textos nuevos con su respectiva polaridad. El beneficio del análisis de sentimiento basado en aprendizaje automático

es que puede manejar nuevas palabras o frases que no estaban presentes en el conjunto de entrenamiento. Sin embargo, se requiere de una gran cantidad de datos etiquetados para entrenar el algoritmo, y el modelo resultante puede ser menos interpretable que un enfoque basado en léxico. (Srivastava, Bharti, & Verma, 2022)

#### 2.3.1.1 REGRESIÓN LOGÍSTICA

La regresión logística es un tipo de modelo estadístico que es usado para la clasificación y análisis predictivo. La regresión logística estima la probabilidad de que un evento ocurra. Como el resultado es una probabilidad, la variable dependiente se limita a valores entre 0 a 1. El modelo se aplica a las probabilidades, la probabilidad de éxito se divide entre la probabilidad de fracaso. Esto se conoce comúnmente como el log de probabilidades, o logaritmo natural de probabilidades. (IBM, 2023)

#### 2.3.1.2 MÁQUINAS DE VECTORES DE SOPORTE

Una máquina de vectores de soporte (SVM) es un modelo de aprendizaje automático supervisado que utiliza algoritmos de clasificación para problemas de clasificación de dos grupos. Después de proporcionar a un modelo SVM conjuntos de datos de entrenamiento etiquetados para cada categoría, pueden categorizar texto nuevo.

En comparación con algoritmos más nuevos como las redes neuronales, tienen dos ventajas principales: mayor velocidad y mejor rendimiento con un número limitado de muestras (miles). Esto hace que el algoritmo sea muy adecuado para problemas de clasificación de texto, donde es común tener acceso a un conjunto de datos de como máximo un par de miles de muestras etiquetadas. (IBM, 2021)

#### 2.3.1.3 APRENDIZAJE SUPERVISADO

La clasificación del sentimiento generalmente es formulada como una clasificación de dos clases: positivo y negativo. La data de entrenamiento y prueba usada son normalmente las reseñas de productos. Ya que, la mayoría de las veces, las reseñas en internet traen consigo un sistema de clasificación (por ejemplo, 1 a 5 estrellas), las clases positivas y negativas son determinadas usando esta clasificación. Por ejemplo, una clasificación de 4 a 5 estrellas es considerada como

una reseña positiva y una reseña con 1 a 2 estrellas es considerada negativa. Una reseña de 3 estrellas podría ser considerada como neutral.

La clasificación del sentimiento es esencialmente un problema de clasificación de texto. Los sistemas de clasificación de texto tradicionales clasifican mayormente documentos en diferentes temas: política, ciencia, deportes, etc. En tales clasificaciones, las palabras relacionadas con el tema son un punto importante. Sin embargo, en la clasificación del sentimiento, sentimientos u opiniones que indican positividad o negatividad son más importantes. Ejemplo: bueno, super, asombroso, horrible, malo, peor, etc.

Ya que es un problema de clasificación de texto, cualquier método de aprendizaje supervisado puede ser aplicado. Por ejemplo, clasificación de Naïve Bayes y máquinas de soporte de vectores (SVM). Como cualquier otra aplicación de aprendizaje supervisada, la clave para la clasificación de sentimientos es la ingeniería de un conjunto de características efectivas (Moreno, Martelo, Corredor, Cifuentes, & Lopez, 2020):

*Términos y sus frecuencias:* Estas características son palabras individuales y la frecuencia con la que aparecen. Es la característica más común en clasificaciones de texto tradicionales. Han mostrado ser bastante efectivo para la clasificación del sentimiento.

*Parte del discurso:* La categoría de cada palabra en una opinión es importante. Palabras en diferentes partes de la opinión pueden ser tratadas diferente. Por ejemplo, está comprobado y tiene lógica pensar que los adjetivos son indicadores importantes en las opiniones. Esto ocasiona que varios investigadores traten a los adjetivos como características especiales de su programa.

*Palabras y frases de sentimientos:* Palabras de sentimientos son usadas para expresar sentimientos positivos o negativos. Por ejemplo: bueno, excelente y asombroso son palabras con una connotación positiva y pésimo, malo, terrible traen consigo una connotación negativa. La mayoría de las palabras de sentimiento son adjetivos y adverbios, pero sustantivos (ejemplo: basura) y verbos (amar / odiar) pueden ser usados también para expresar sentimientos. Además de palabras individuales, frases pueden ser usadas para expresar sentimientos: “mejor me quedo sin comer antes de probar eso”.

#### 2.3.1.4 APRENDIZAJE NO SUPERVISADO

Ya que las palabras de sentimientos son usualmente el factor dominante para la clasificación del sentimiento no es difícil pensar que las palabras o frases de sentimientos pueden ser usadas para la clasificación de sentimientos en una manera no supervisada. Una técnica es desarrollar la clasificación de sentimiento basada en patrones sintácticos fijos que es probable que se utilicen para expresar opiniones. Los patrones sintácticos se componen a base de etiquetas de parte del discurso (Peter, 2002). Este método consiste en tres pasos:

*Paso 1:* Dos palabras consecutivas son extraídas si sus etiquetas de parte del discurso son conforme a un patrón ya establecido. Por ejemplo, en la oración “Este piano produce un sonido hermoso”, “sonido hermoso” es extraído como el primer patrón. Los sustantivos y verbos actúan como el contexto y estos pueden ocasionar que el sentimiento clasificado sea diferente.

*Paso 2:* Se genera un valor de orientación del sentimiento de las frases extraídas usando una fórmula que mide el grado estadístico de dependencia entre las dos palabras.

*Paso 3:* Con el valor obtenido del paso dos, el algoritmo calcula un promedio y en base a este se decide si la connotación es positiva o negativa.

#### 2.3.2 METODOLOGÍAS DESARROLLADAS POR OTROS INVESTIGADORES O EXPERTOS

En otras investigaciones, el análisis de sentimiento se lleva a cabo sin dividir la data en los aspectos latentes encontrados en el conjunto de datos. Esta pérdida de contexto ocasiona que ciertas palabras ambiguas no puedan tener su polaridad real clasificada correctamente. Como se mencionó anteriormente, el *contexto* es una parte fundamental para identificar el sentimiento correcto de un documento.

En el trabajo hecho por (Tumasjan, Sprenger, Sandner, & Welpe, 2010), se examinaron 104,003 tweets políticos que fueron publicados en la misma red social entre agosto 13, 2019 a septiembre 19, 2019. Se filtraron buscando palabras claves relacionadas con políticos de interés. Para obtener el sentimiento de estos tweets, se usó un software especializado de análisis de texto al cual se le alimentó un diccionario altamente personalizado. El software calcula lo que llaman dimensiones psicológicas basadas en palabras que fueron empíricamente categorizadas en

diferentes tipos. Por ejemplo, palabras como ‘tal vez’, ‘quizás’ y ‘supongo’ están categorizadas como de la dimensión ‘tentativos’. El trabajo realizó 12 diferentes dimensiones relacionadas con emociones como ser tristeza, ansiedad, felicidad, enojo, etc.

Uno de los mayores retos para este tipo de metodologías en el análisis de sentimiento es que si se encuentran palabras que no están ingresadas en el diccionario en uso, la identificación de las categorías definidas falla y se pierde el sentimiento detrás de ellas. También dicho método no obtiene el suficiente contexto necesario para poder identificar aquellas palabras cuya polaridad varía según el mismo.

En el trabajo realizado por (Nurifan, Sarno, & Sungkono, 2019), se realiza un análisis de sentimiento basado en aspectos sobre las reseñas de restaurantes. Ellos proponen un proceso de cuatro etapas que incluyen Extracción de los términos de Aspectos, Extracción de palabras claves de los Aspectos, categorización de aspectos y análisis de sentimiento. Para la extracción de los términos, se utilizan diferentes métodos de modelado de aspectos en conjunto con un extractor de sustantivos y filtro de términos usando etiquetas de parte de discursos. En el paso de extracción de palabras claves de términos, se utiliza un método semi supervisado para extraer todas las palabras claves de los aspectos usando TC-ICF y data preprocesada para disminuir los esfuerzos de manualmente extraer las palabras claves una a una.

Al momento de la categorización de aspectos, se propone un híbrido del método EIMo-Wikipedia para determinar los vectores para los términos de aspectos y palabras claves basados en el contexto.

Para terminar, el trabajo propone usar un híbrido de expansión de opinión de Lexicon-SentiCircle para cambiar la polaridad de palabras en el léxico de opinión basadas en los aspectos. Las metodologías mostradas en ese papel pueden mejorar la precisión del análisis de sentimiento basado en aspectos para reseñas de restaurantes.

## **CAPITULO III. METODOLOGÍA DE LA INVESTIGACIÓN**

### **3.1 CONGRUENCIA METODOLÓGICA**

La congruencia metodológica ayuda en las investigaciones a que se realicen con un proceso de lógica de las etapas que se están investigando. La congruencia metodológica debe mostrar la relación existente entre el problema planteado y la metodología propuesta para encontrar variables que lleven a encontrar una solución al problema planteado. En el caso de esta investigación el enfoque es cuantitativo con un alcance descriptivo. Esto debido a que se busca describir el estado actual de las variables seleccionadas sobre un restaurante, proporcionando información sistemática sobre el peso que cada variable influye sobre el éxito del negocio.

#### **3.1.1 MATRIZ METODOLÓGICA**

Con el objetivo de poder conceptualizar todas las variables de la investigación se realiza un cuadro para el ordenamiento de la información que va a ser necesaria para el entendimiento del proyecto.

A continuación, se muestra la matriz de congruencia metodológica en la Tabla 2, en este cuadro se muestra la relación que debe existir entre el planteamiento del problema, las preguntas de investigación y los objetivos generales y específicos; la congruencia entre las variables dependientes e independientes. En resumen, detalla el planteamiento del problema y la ruta esquematizada de cada uno de sus componentes.

Tabla 2 Matriz de congruencia metodológica

Formulación del Problema	Objetivo General	Objetivos específicos	Preguntas de Investigación	Variable Dependiente	Variables Independientes
¿Es posible realizar un análisis de sentimiento con la información de las reseñas en Google Maps de una cadena de restaurantes en Honduras?	Realizar mediante machine learning un análisis del sentimiento y categorización de la retroalimentación brindada por clientes de una cadena de restaurantes de comida rápida en Google Maps.	Categorizar cada reseña identificada en cuatro aspectos culinarios predefinidos	¿Cuáles son los aspectos más mencionados en las reseñas del restaurante?	Local	Índice de Sentimiento
		Identificar mediante el análisis del sentimiento la polaridad de los aspectos o temas definidos en las reseñas de los comensales	¿Cuál es la polaridad de los aspectos según las reseñas del restaurante?	Precio	
				Servicio	
Determinar las oportunidades de mejora en los aspectos según el análisis del sentimiento de los comentarios dejados por los clientes	¿Cuáles son las oportunidades de mejora identificadas en el negocio?	Comida			

Fuente: Propia

### 3.1.2 OPERACIONALIZACIÓN DE LAS VARIABLES

Operacionalización de las variables es un proceso que obliga a realizar una definición conceptual de las variables para romper el concepto difuso que ella engloba y así darle sentido concreto dentro de la investigación, luego se procesa a realizar la definición operacional para identificar los indicadores que permitirán realizar su medición de forma empírica y cuantitativa, al igual que cualitativamente llegado el caso.

*Entorno físico:* Al momento de establecer un negocio, el local es sumamente importante. La ubicación (si está en una zona de alta plusvalía o alto tránsito). Si tiene empresas como maquila o centros de estudio como universidades cerca. Todas estas características ayudarán a establecer cuál debe ser el público objetivo. Si está cerca de una universidad, vale la pena pensar en tener un local amplio con bastantes mesas y sillas para que los estudiantes puedan llegar y relajarse.

Siguiendo el último punto, si el local se ubica cerca de una universidad, tal vez sea necesario crear un ambiente de relajación en el cual los estudiantes puedan sentarse a consumir su producto y estudiar y/o trabajar en tareas. Si está cerca de un centro de convenciones o una zona viva, es necesario generar un ambiente vivo.

*Precio:* El precio de los productos y/o servicios es importante. Son influenciados por el local, pago de salarios, pago de insumos, pago de servicios básicos. Es importante tomar en cuenta la zona en que se encuentra el local, si es una zona de alta plusvalía entonces tener precios un poco alto puede ser posible.

*Servicio:* El servicio es importante para un negocio. Tratar al cliente con respeto y estar al tanto a las necesidades de él o ella puede ser clave para la retención de este. Muchas veces, gestos tan sencillos como memorizar el nombre de un cliente puede generar un sentido de lealtad en el hacia el negocio.

*Comida:* El producto a vender en el negocio. Si el producto es de mala calidad y no tiene relación con el precio de este, el negocio fallará.

Tabla 3 Operacionalización de las variables

Variable	Definición		Dimensiones	Indicador
	Conceptual	Operacional		
Comida	Se conoce como comida al conjunto de sustancias alimenticias que se comen y se beben para subsistir. (Pérez Porto & Merino, 2021)	Alimento / Bebida ofrecido en el restaurante	Calidad	Ingredientes frescos
				Servido a temperatura adecuada
			Sabor	Nutrición
Buen sabor				
Cantidad	Tamaño de la porción			
Servicio	Con origen en el término latino <i>servitium</i> , la palabra servicio define a la actividad y consecuencia de servir. (Pérez Porto & Merino, 2021)	El servicio provisto por parte de los empleados del lugar a los comensales	Exactitud	Se recibe exactamente lo que se ordena
				Disponibilidad de lo ofrecido en el menú
			Confiabilidad	Los empleados conocen el menú
Los empleados están dispuestos a ayudar				
Reacción	Rapidez en el servicio			
	Reacción a solicitudes del cliente			
Entorno físico	Las condiciones (físicas, económicas, culturales, etc.) de un lugar, un grupo o una época. (Pérez Porto & Merino, 2021)	Las condiciones físicas del local y el ambiente construido por decoración y/o música	Ambiente	Higiene del lugar
				Atmosfera del lugar
			Local	Espacio disponible en mesa
Decoración del lugar				
Precio	Es el valor monetario que se le asigna a algo. (Pérez Porto & Merino, 2021)	Valor monetario de los productos y/o servicios disponibles en el lugar	Valor monetario	Precios razonables
				Buen valor por el dinero
			Valor percibido	Valor general de la experiencia

Fuente: Propia

### 3.1.3 HIPÓTESIS

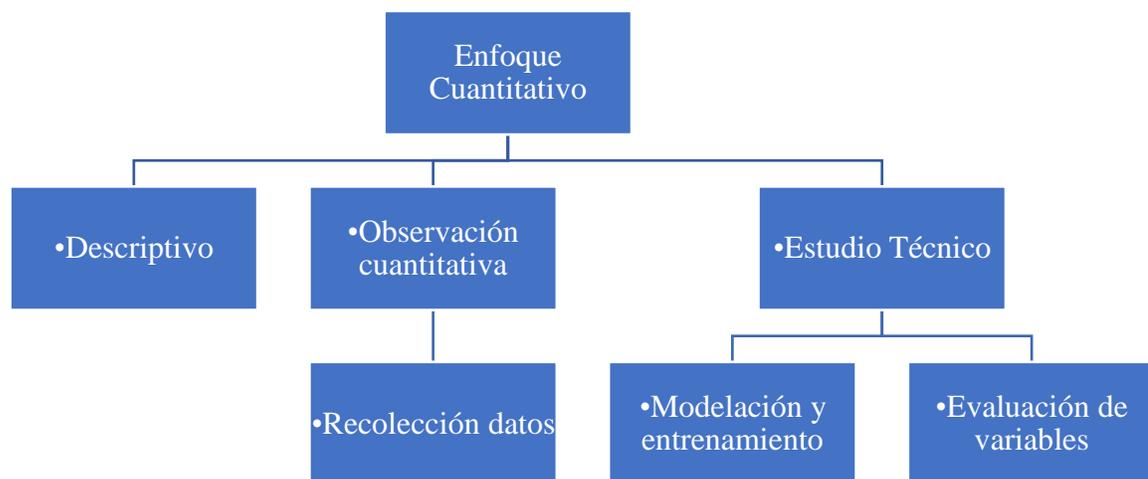
No definidas para el trabajo actual considerando un alcance descriptivo de la investigación.

## 3.2 ENFOQUE Y MÉTODOS

Las investigaciones cuantitativas recolectan y analizan información numérica. Es ideal para identificar tendencias y promedios, hacer predicciones, comprobar relaciones y generalizar resultados para grandes poblaciones. Este método se usa mayormente en ciencias naturales y sociales como psicología, economía, social, marketing, entre otras. Es una investigación sistemática de un fenómeno realizada mediante la recolección de data cuantificable y performance análisis usando técnicas estadísticas, matemáticas y computacionales. Esta investigación obtiene información de clientes existentes y potenciales usando métodos de muestreo y enviando encuestas en línea o haciendo minería de datos en redes sociales. (Ortega, n.d.)

A continuación, se define el esquema metodológico de investigación a utilizar:

*Ilustración 1 Esquema metodológico de investigación*



Fuente: Propia

### **3.3 DISEÑO DE LA INVESTIGACIÓN**

#### **3.3.1 POBLACIÓN**

En esta investigación, se tendrá como población las reseñas vertidas en Google Maps de la cadena de restaurantes Denny's en Honduras. En base a lo anterior, se encontraron un total de 7,495 reseñas únicas esparcidas en los restaurantes de Denny's que se ubican en Honduras.

#### **3.3.2 MUESTRA**

El análisis por desarrollarse en esta investigación define que esta población se va a abordar mediante la técnica de observación de datos obtenido de minería de texto. Por lo tanto, la muestra no probabilística que se obtiene es el análisis de los datos obtenidos de la minería llevada a cabo en un periodo de tiempo de 1 año.

#### **3.3.3 TÉCNICAS DE MUESTREO**

Mediante el uso de la plataforma Apify, que permite obtener información de internet mediante web scraping, se obtendrán las reseñas de los restaurantes de Denny's en Honduras que se encuentran en Google Maps. Apify permite desarrollar, construir y ejecutar trabajos de software de larga duración llamados Actores, que son extremadamente útiles para tareas de automatización web. (Čurn, 2019).

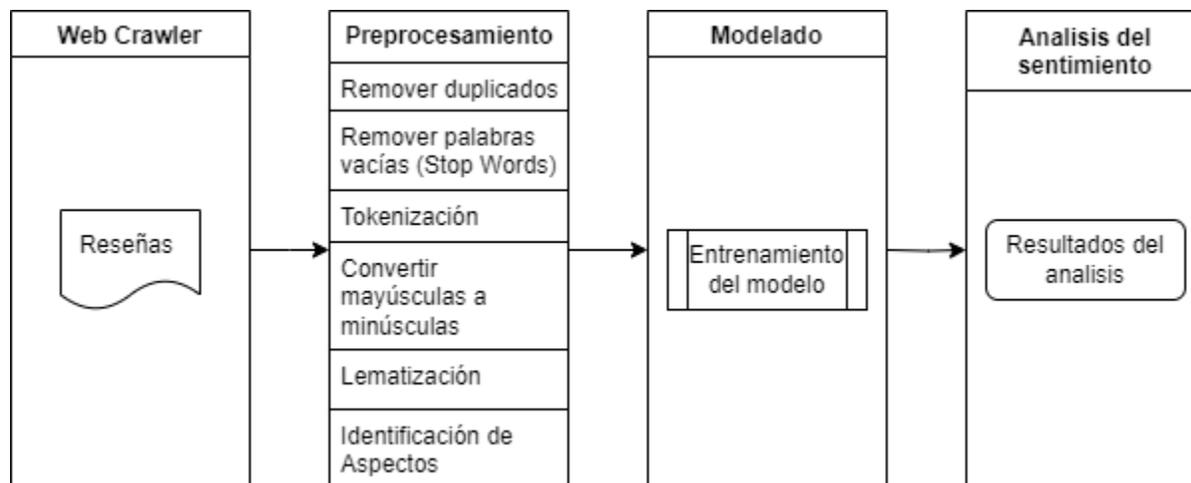
A demás de desarrollar actores, la plataforma permite hacer uso de soluciones creadas por otros usuarios. En esta investigación se hizo uso del Actor que permite bajar reseñas de un lugar específico de Google Maps al proveerle la dirección URL de este.

### 3.4 TÉCNICAS, INSTRUMENTOS Y PROCEDIMIENTOS APLICADOS

#### 3.4.1 TÉCNICAS

En este trabajo, proponemos un marco de análisis de sentimientos basado en aprendizaje profundo para reseñas de la cadena de restaurantes *Denny's* en Honduras. El marco de investigación se muestra en la Ilustración 2 con cuatro componentes.

Ilustración 2 Marco de investigación



#### 3.4.2 PROCEDIMIENTOS

##### 3.4.2.1 PREPROCESAMIENTO

El preprocesamiento es un paso inicial cuyo objetivo es preparar el conjunto de datos para que se puedan transformar a una manera más estructurada, asegurando que la información a procesarse en los modelados sea data íntegra y de calidad.

Este paso incluye los siguientes:

- Se convierte todo el texto a minúsculas
- Se eliminan caracteres que no sean alfabéticos
- Se eliminan registros duplicados y/o nulos
- Se normaliza las palabras, eliminando las palabras con tilde y reemplazándolas a su versión sin tilde

- Se eliminan las palabras vacías o *stopwords*. Siendo estos usualmente pronombres, adverbios, preposiciones, etc.
- Se lematizan las palabras. El lema es la representante de todas las formas de una misma palabra. Por ejemplo, decir es el lema de dije, pero también de diré o dijéramos; mesa es el lema de mesas.
- Se encuentran el PoS tag de las palabras, una técnica fundamental en NLP que consiste en etiquetar cada palabra de un documento en su correspondiente categoría gramatical.
- Se revisa si el conjunto de datos final es balanceado o no al tener una cantidad similar de reseñas con una connotación positiva y negativa.

A continuación, detallamos un ejemplo en la Ilustración 3 del resultado del preprocesamiento ejecutado en la data. Se quitaron los caracteres que no sean alfabéticos, se normaliza las palabras, se eliminan las palabras vacías y se lematizan.

*Ilustración 3 Ejemplo de reseña procesada*

<b>original</b>	Excelente servicio y muy buena atención y lo mejor que si en verdad funciona 24/7
<b>procesada</b>	excelente servicio atencion mejor

Definimos anteriormente que el análisis de sentimientos y clasificación es esencialmente un problema de clasificación de texto. La clasificación de texto se realiza haciendo uso de Part-Of-Speech Tagging (PoS Tag), una técnica fundamental en procesamiento de lenguaje natural, consistiendo en etiquetar cada palabra encontrada en una oración o documento dado en su correspondiente categoría gramatical. En esencia, se busca categorizar las palabras en adverbios, sustantivos, pronombres, acrónimos, etc. Sumando a esto, una técnica más es la identificación de palabras relacionadas con el tema de investigación, en este caso un ejemplo puede ser: bueno, delicioso, asqueroso, barato o caro, insípido, etc. Esto ayuda al modelo poder lematizar, o sea el proceso de agrupar palabras para que puedan ser analizadas como un solo ítem. Un ejemplo de un análisis PoS es el siguiente: *Mi nombre es Pablo y vivo en Honduras*. Se visualiza en la Tabla 4.

Tabla 4 Análisis PoS de una oración

Token	PoS
Mi	DET (determinante, palabras que modifican a sustantivos)
Nombre	NOUN (sustantivo, denotan personas, animales, cosas, etc.)
Es	AUX (auxiliar, verbos sin un significado pleno, por eso no se clasifican como tal)
Y	CCONJ (conjunción, palabra que une otras)
Vivo	VERB (verbo)
En	ADP (preposiciones y posposiciones)
Honduras	PROPN (pronombre propio, nombre de un lugar, individuo u objeto específico)
.	PUNCT (puntuación, usado para delimita)

En este trabajo se utilizará la librería Spacy de Python, es una librería para procesamiento de texto avanzado de lenguaje natural. Esta incluye modelo preentrenados para la predicción de etiquetas PoS, dependencias sintácticas o entidades propias. Spacy cuenta con diferentes librerías para varios lenguajes, entre ellos el castellano. El modelo que se utilizara para el entrenamiento es *es\_core\_news-md*. Es una red neuronal convolucional entrenada con textos periodísticos.

El propósito de lematizar nuestras reseñas es retornar las palabras de estas a su base. El lematizado de Spacy es responsable de agrupar diferentes grupos de formas de palabras en su forma base, todas teniendo en esencia el mismo significado.

### 3.4.2.2 MODELADO DE ASPECTOS (LDA)

LDA es un modelo generativo no supervisado que categoriza las palabras que aparecen en documentos en grupos referidos como aspectos o temas. La idea básica del LDA es que un documento consiste en uno o varios aspectos o temas latentes donde cada uno consiste en una distribución de palabras que los identifica. En el contexto del trabajo mencionado, se utilizó LDA con los parámetros 'max\_df' y 'min\_df', que eliminan del corpus términos que aparecen con demasiada o poca frecuencia. En este caso, los parámetros que arrojaron mejores resultados fueron eliminar términos que aparecieron en más del 70% de las reseñas y términos que aparecieron en menos de 5 reseñas. El parámetro 'n\_components' se configuró en 4 para clasificar las reseñas en los temas esperados: precio, lugar, servicio y comida. Podemos ver el código con los parámetros en la Ilustración 4.

```
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.decomposition import LatentDirichletAllocation

count_vect = CountVectorizer(max_df=0.7, min_df=5, stop_words=['no'])
doc_term_matrix = count_vect.fit_transform(reviews_datasets['cleaned_reviews'].values.astype('U'))
print(doc_term_matrix.shape)

LDA = LatentDirichletAllocation(n_components=4, random_state=42, learning_decay=0.7)
LDA.fit(doc_term_matrix)
```

Ilustración 4 Código Python del LDA

Con los parámetros anteriormente mencionados, se logra obtener 4 temas latentes encontrados en todo el corpus. En la Ilustración 5 se detallan estos temas, con las palabras de mayor peso ordenadas de mayor a menor.

```
Topic #0: ['servir', 'tardar', 'malo', 'personal', 'servicio', 'cliente', 'lugar', 'mesero', 'atender', 'comer']
Topic #1: ['hamburguesa', 'esperar', 'restaurante', 'tiempo', 'excelente', 'cafe', 'super', 'delicioso', 'desayuno', 'rico']
Topic #2: ['rico', 'rapido', 'servicio', 'calidad', 'delicios', 'hora', 'mejor', 'precio', 'lugar', 'comida']
Topic #3: ['agradable', 'menu', 'gustar', 'lugar', 'encantar', 'ambiente', 'servicio', 'comida', 'atencion', 'excelente']
```

Ilustración 5 Resultados del LDA con parámetros establecidos

Podemos observar que el Tema #0 contiene palabras relacionados con el *servicio* como ser ‘servir’, ‘tardar’, ‘malo’. En la Tabla 5 podemos ver un ejemplo de una reseña que según la ponderación del algoritmo se definió trataba sobre el aspecto de *servicio*.

Tabla 5 Categorización del algoritmo LDA a Servicio

reviews	cleaned_reviews	Servicio	Comida	Precio	Lugar	dominant_topic
En el del bulevar Juan Pablo, tienen un pésimo servicio al cliente. Puede estar vacío y lo que hacen es estar sentados todos los empleados en una mesa al fondo y uno tiene que ir a pedirle las cosas hasta 3 veces.	bulevar juan pablo pesimo servicio cliente vacio sentado empleado mesa fondo pedir el cosa	0.95	0.02	0.02	0.02	0

El Tema #1 contiene palabras relacionadas con la *comida* como ser ‘hamburguesa’, ‘esperar’, ‘restaurante’, ‘café’, ‘delicioso’. En la Tabla 6 podemos ver un ejemplo de una reseña que según la ponderación del algoritmo se definió trataba sobre el aspecto de *comida*. La misma reseña tuvo una alta ponderación en *servicio*. Como era de esperarse, una reseña puede aludir a más de un tema. El algoritmo hace un buen trabajo identificando el peso latente de aspecto en las reseñas.

Tabla 6 Categorización del algoritmo LDA a Comida

reviews	cleaned_reviews	Servicio	Comida	Precio	Lugar	dominant_topic
aquí todo es rico, lo malo es que pasa lleno casi siempre y toca esperar.	rico malo pasar lleno tocar esperar	0.45	0.48	0.04	0.04	1

El Tema #2 es un poco más ambiguo de identificar, pero podemos ver que trae en consigo la palabra ‘precio’ por lo que podemos definir que está relacionado con el *precio*. En la Tabla 7, podemos ver el siguiente ejemplo en donde la reseña tuvo el mayor peso en el aspecto del *precio*, teniendo como segundo aspecto *comida*.

Tabla 7 Categorización del algoritmo LDA a Precio

<b>reviews</b>	<b>cleaned_reviews</b>	<b>Servicio</b>	<b>Comida</b>	<b>Precio</b>	<b>Lugar</b>	<b>dominant_topic</b>
Buen servicio pero creo que la comida puede mejorar. Precios elevados.	servicio crear comida mejorar precio elevado	0.35	0.04	0.58	0.04	2

El Tema #3 trata con el lugar, podemos ver que hay palabras relacionadas como ser ‘agradable’, ‘lugar’ y ‘ambiente’. En la Tabla 8 vemos un ejemplo donde el algoritmo identifico el aspecto de *lugar* y seguidamente lo identifico como *servicio*.

Tabla 8 Categorización del algoritmo LDA a Lugar

<b>reviews</b>	<b>cleaned_reviews</b>	<b>Servicio</b>	<b>Comida</b>	<b>Precio</b>	<b>Lugar</b>	<b>dominant_topic</b>
Es un lugar con muy buen ambiente, con personal dispuesto a complacer a sus clientes y con una buena variedad de comida	lugar ambiente personal dispuesto complacer cliente variedad comida	0.46	0.03	0.03	0.47	3

### 3.4.3 INSTRUMENTOS

Con este conjunto de datos, una vez se haya entrenado el modelo se aplicará una medida de comparación para definir el método más certero. El cuadro de la matriz de confusión se puede encontrar en Anexo 3. Este cuadro permitirá demostrar las reseñas correctamente clasificadas como positivas (Verdadero Positivo), el número de negativos clasificados como positivos (Falso Positivo), el número de positivos clasificados como negativos (Falso Negativo) y el número de negativos correctamente clasificados (Verdadero Negativo). Para comparar los métodos de evaluación, se considera las mediciones siguientes con una unidad basada en porcentajes:

*Exactitud (Accuracy):* se refiere a lo certeza del modelo en clasificar correctamente reseñas. Se formula así:  $(VP + VN) / (VP + FP + FN + VN) \%$

*Precisión:* Se refiere a la dispersión del conjunto de valores obtenidos a partir de mediciones repetidas de una magnitud. Cuanto menor es la dispersión mayor la precisión. Se formula así:  $VP / (VP + FP) \%$

*Sensibilidad:* También se conoce como tasa de verdaderos positivos. Es la proporción de casos positivos que fueron correctamente identificadas por el algoritmo. Se formula así:  $VP / (VP + FN)$

*F1 Score:* Esta es otra métrica muy empleada porque nos resume la precisión y sensibilidad en una sola métrica. Por ello es de gran utilidad cuando la distribución de las clases es desigual, por ejemplo, cuando el número de pacientes con una condición es del 15% y el otro es 85%, lo que en el campo de la salud es bastante común. (Barrios, 2019).

Se calcula en porcentajes y se formula de la siguiente manera vista en la Ilustración 6.

*Ilustración 6 Formula del Puntaje F1*

$$\text{PuntajeF1} = \frac{2 * \text{Precision} * \text{Sensibilidad}}{\text{Precision} + \text{Sensibilidad}}$$

Fuente: (Barrios, 2019)

Los resultados de la métrica anterior serán mostrados como lo indica el Anexo 4. Se propone también dar un resultado resumen general de las fuentes identificadas. Este resumen

pretende darle a negocio una vista superficial de la polaridad actual que sus redes sociales poseen al momento de realizar el análisis. Se puede ver el cuadro en Anexo 2.

### **3.5 FUENTES DE INFORMACIÓN**

#### **3.5.1 FUENTES PRIMARIAS**

Como fuente en esta investigación, se usará las reseñas u opiniones dejadas por usuarios que se asumen fueron consumidores del negocio escogido. La fuente primaria por escoger en este caso sería Google Maps Reviews.

#### **3.5.2 FUENTES SECUNDARIAS**

No hay fuentes secundarias para el presente trabajo.

## CAPITULO IV. RESULTADOS Y ANÁLISIS

### 4.1 DESCRIPCIÓN DEL RESTAURANTE

El estudio se basa en las reseñas dejadas en Google Maps por usuarios en las diferentes sucursales de la cadena de restaurantes *Denny's* ubicados en Honduras. *Denny's* es una cadena de restaurantes estilo comedor estadounidense con servicio de mesa. Opera más de 1,700 restaurantes en muchos países. Actualmente en Honduras se encuentran seis restaurantes, 3 en San Pedro Sula, Cortes y 3 en Tegucigalpa.

### 4.2 RECOLECCIÓN DE DATOS

Haciendo uso de la plataforma *Apify* que automatiza el proceso de recolección de datos en la web, se obtuvo las reseñas dejadas por los usuarios en las 6 diferentes sucursales de *Denny's* y se procedió a crear 6 diferentes conjuntos de datos. En la recolecta inicial de datos no se definió un límite de tiempo, se eligió traer todas las reseñas sin importar la fecha en que se dejaron, esto por el motivo de tener más datos para el entrenamiento del modelo de clasificación de sentimiento. Para motivos de los resultados finales, se usarán las reseñas que se dejaron en un lapso de 1 año. Se detalla abajo una descripción en la Tabla 9 de cada conjunto:

Tabla 9 Conjunto de datos obtenidos

Nombre Dataset	Lugar	Cantidad reseñas
dataset_GoogleMapReviews_01	SPS Denny's Blvr Morazán	1868
dataset_GoogleMapReviews_02	TGU Denny's Blvr Juan Pablo	1776
dataset_GoogleMapReviews_03	SPS Denny's Jardines	1210
dataset_GoogleMapReviews_04	TGU Denny's Toncontín	1062
dataset_GoogleMapReviews_05	TGU Denny's Los Próceres	834
dataset_GoogleMapReviews_06	SPS Denny's Blvr Este	745

Como resultado del preprocesamiento de los conjuntos de datos se obtuvo 7,495 diferentes reseñas, cada una con su respectiva calificación de estrellas establecida por los usuarios.





## 4.4 RESULTADOS Y ANÁLISIS DE LAS TÉCNICAS APLICADAS

Después del preprocesamiento del conjunto de datos, procedemos a ver la distribución de

### 4.3.1 REGRESIÓN LOGÍSTICA (LR)

Al usar la regresión logística con los documentos ya procesados, se encontró que el modelo tuvo un F1-Score del 95.56%. La matriz de confusión mostrada en la Ilustración 11 nos demuestra que el algoritmo de regresión logística logro identificar correctamente 1835 reseñas con una polaridad negativa pero fallo al identificar 91 reseñas con una polaridad positiva cuando eran negativas. Tenemos un resultado similar con las reseñas positivas, el algoritmo identifico 1846 correctamente como positivas y fallo clasificando negativamente 80 reseñas que eran positivas.

Ilustración 11 Resultados matriz de confusión de LR

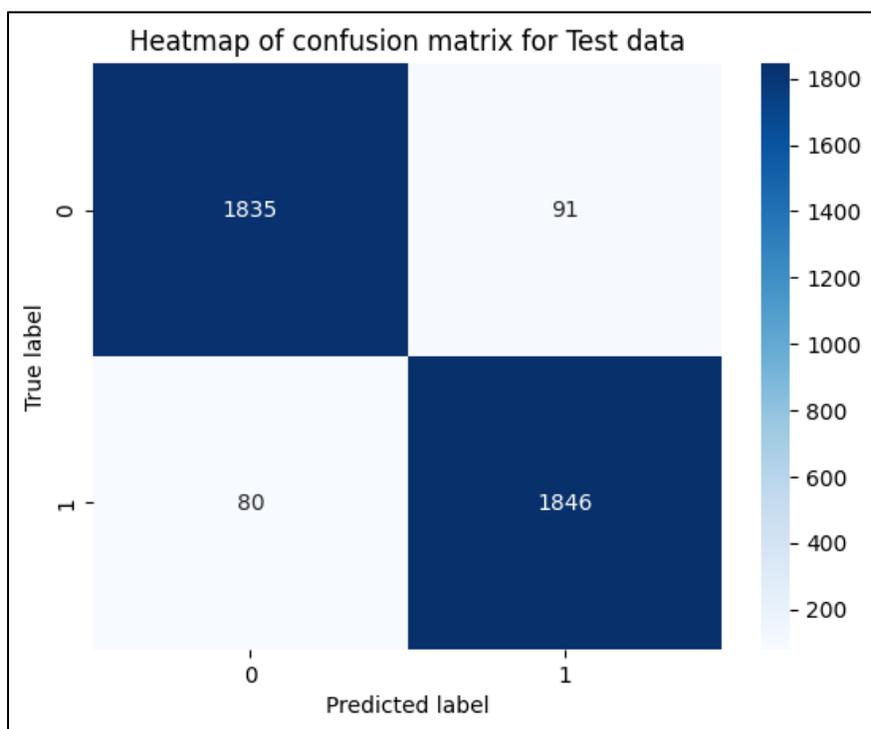


Ilustración 12 Resultados de medición de LR

	precision	recall	f1-score	support
0	0.9582	0.9528	0.9555	1926
1	0.9530	0.9585	0.9557	1926
accuracy			0.9556	3852
macro avg	0.9556	0.9556	0.9556	3852
weighted avg	0.9556	0.9556	0.9556	3852

### 4.3.2 MÁQUINAS DE VECTORES DE SOPORTE (SVM)

Al usar el SVM con los documentos ya procesados, se encontró que el modelo tuvo un F1-Score del 95.09%. La matriz de confusión mostrada en la Ilustración 13 nos demuestra que el algoritmo de máquinas de vectores de soporte logro identificar correctamente 1783 reseñas con una polaridad negativa pero fallo al identificar 143 reseñas con una polaridad positiva cuando eran negativas. Tenemos un mejor resultado con las reseñas positivas, el algoritmo identifico 1880 correctamente como positivas y fallo clasificando negativamente 46 reseñas que eran positivas.

Ilustración 13 Resultados matriz de confusión de SVM

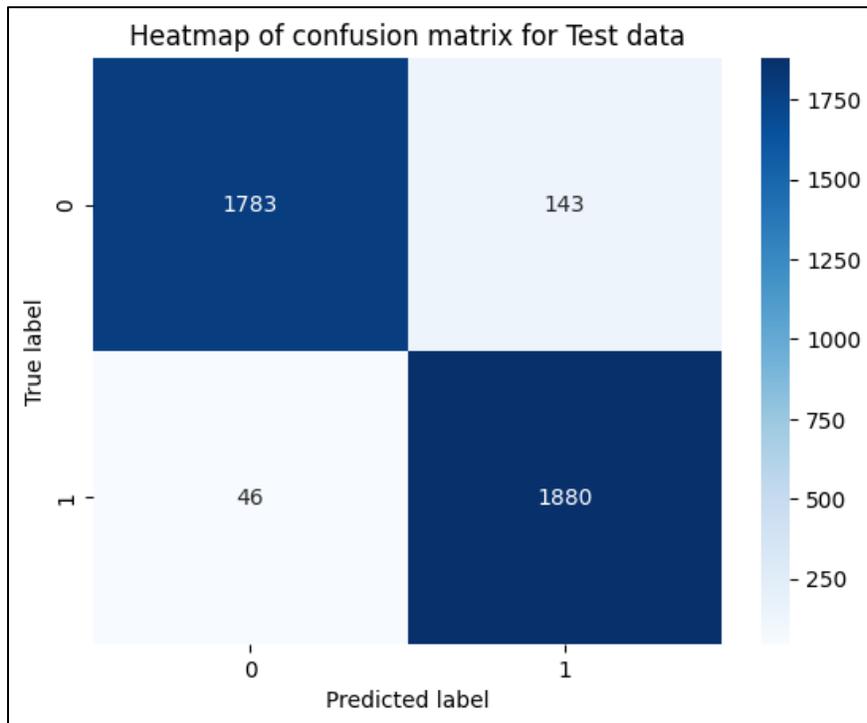


Ilustración 14 Resultados de medición de SVM

	precision	recall	f1-score	support
0	0.9748	0.9258	0.9497	1926
1	0.9293	0.9761	0.9521	1926
accuracy			0.9509	3852
macro avg	0.9521	0.9509	0.9509	3852
weighted avg	0.9521	0.9509	0.9509	3852

### 4.3.3 REGRESIÓN LOGÍSTICA CON AJUSTE DE PARAMETROS (LR-ADJ)

Al usar la regresión logística con un ajuste en sus parámetros con los documentos ya procesados, se encontró que el modelo tuvo un F1-Score del 94.99%. La matriz de confusión mostrada en la Ilustración 15 nos demuestra que el algoritmo de regresión logística ajustado logro identificar correctamente 1793 reseñas con una polaridad negativa pero fallo al identificar 133 reseñas con una polaridad positiva cuando eran negativas. Tenemos un mejor resultado con las reseñas positivas, el algoritmo identifico 1866 correctamente como positivas y fallo clasificando negativamente 60 reseñas que eran positivas.

Ilustración 15 Resultados matriz de confusión de LR-ADJ

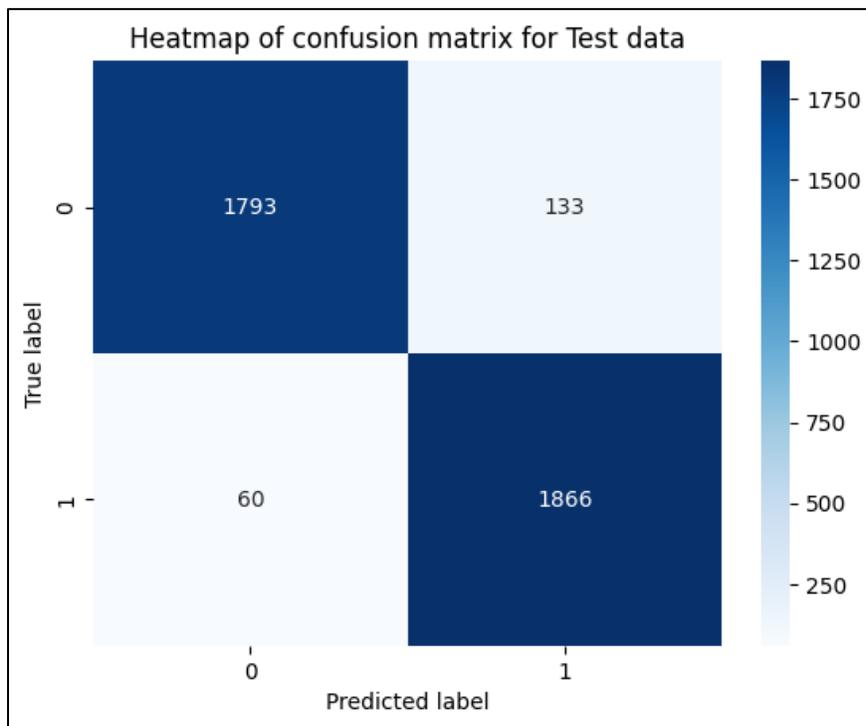


Ilustración 16 Ilustración 8 Resultados de medición de LR-ADJ

	precision	recall	f1-score	support
0	0.9676	0.9309	0.9489	1926
1	0.9335	0.9688	0.9508	1926
accuracy			0.9499	3852
macro avg	0.9505	0.9499	0.9499	3852
weighted avg	0.9505	0.9499	0.9499	3852

## 4.5 RESULTADOS

A continuación, en la Tabla 10 se indican los resultados del análisis de sentimiento en base a los 4 aspectos de lugar, precio, servicio y comida. Los resultados se calcularon con las reseñas dejadas a través de los 6 restaurantes de *Denny's* dejados en un lapso desde enero 2022 hasta septiembre 2023, dejando un total de 1296 reseñas únicas.

Tabla 10 Resultados del modelo

Análisis de sentimiento por Aspecto			
Aspecto	Sentimiento	Total	%
Lugar	Positivo	499	93%
	Negativo	40	7%
Precio	Positivo	351	90%
	Negativo	37	10%
Servicio	Positivo	95	64%
	Negativo	54	36%
Comida	Positivo	142	65%
	Negativo	78	35%

Después de categorizar los aspectos usando el método LDA, es posible ahora realizar un análisis de sentimientos usando el modelo de regresión logística, al tener estos los mejores resultados del experimento.

Nos damos cuenta de que, de las 1296 reseñas, podemos ver que los comensales califican el 'Lugar', el restaurante como tal, de una manera bastante positiva con un 93% de positividad. El aspecto del 'Precio' le sigue, teniendo este un 90% de positividad. El 'Servicio' y 'Comida' tienen un resultado bastante similar, siendo estos mayormente positivos con un 64.5%, pero teniendo el porcentaje de negatividad más alto con un promedio de 35.5%.

Un ejemplo de las reseñas que afectan negativamente el aspecto de Servicio es la siguiente:

Tabla 11 Ejemplo de reseña procesada y clasificada

reviews	sentiment	cleaned_reviews	Servicio	Comida	Precio	Lugar
En el Juan Pablo II La atención muy pésima al final no familia tomó la decisión de irnos ya que él mesero se presentó y no regreso por el pedido.	0	juan pablo ii atencion pesimo no familia tomar decision ir yo el mesero presentar no regreso pedido	0.95	0.02	0.02	0.02

Según el modelado de aspectos, la reseña trata mayormente sobre servicio con un 95% de similitud y trae consigo una polaridad negativa.

## CAPITULO V. CONCLUSIONES Y RECOMENDACIONES

### 5.1 CONCLUSIONES

- Después de aplicar el modelado de aspectos, procedimos con la clasificación del sentimiento mediante el modelo de regresión logística (LR). Este modelo obtuvo el mejor resultado con un 95.56%. Los modelos de SVM y LR-ADJ obtuvieron un F1-Score de 95.09% y 94.99% respectivamente. Al obtener los sentimientos, el resultado obtenido indica que el aspecto mejor calificado es el 'Lugar' con un 93% de positividad, siguiéndole el 'Precio' (90%), 'Comida' (65%) y 'Servicio' (64%).
- Según el conjunto de datos, las reseñas obtenidas de Google Maps por medio de Apify dejadas por los usuarios hablan mayormente sobre el 'Lugar'. Esto quiere decir, que los comensales tienen una opinión bastante favorable sobre el ambiente que el restaurante Denny's ofrece. Son 539 (41%) de las reseñas que según el modelado de aspectos hablan sobre el 'Lugar'. De esa cantidad, 499 (93%) son reseñas positivas.
- En este trabajo, se logró obtener un modelo de procesamiento del lenguaje natural realizado en Python sobre el análisis de sentimiento en reseñas obtenidas de un restaurante. Se comprobó la factibilidad de construir un algoritmo propio de reconocimiento de sentimientos a un nivel de precisión de 95.56%.
- Según los resultados, los dos aspectos con una polaridad negativa en las reseñas obtenidas son 'Servicio' (36%) y 'Comida' (35%). En base a esto, se puede hacer un análisis más exhaustivo de estos dos aspectos en los restaurantes de Denny's para verificar la causa raíz de la inconformidad de los clientes con ellos.
- Debido a la naturaleza del trabajo, para poder ser aplicable este modelo en diferentes restaurantes es necesario que estos tengan una presencia activa en la red para poder obtener una buena cantidad de retroalimentación dejada por los comensales.

## 5.2 RECOMENDACIONES

- Se recomienda a negocios con una presencia activa en redes a hacer el uso de métodos de análisis de sentimientos en las interacciones que obtienen en las redes sociales. Esta es una información valiosa que ayudaría a identificar tendencia y tomar decisiones de negocio certeras que incrementen el nivel de lealtad en sus clientes y así poder crecer las ganancias.
- El objetivo final del modelo desarrollado es clasificar cada reseña e identificar la polaridad. Para motivos del entrenamiento inicial, se obtuvo reseñas ya con una clasificación en base a estrellas que indicaba el sentimiento. Con el modelo entrenado, esta parte no ya no es necesaria. Se puede alimentar comentarios sin clasificación y el modelo identificara la polaridad y el aspecto asignado.
- El modelo hace uso de la clasificación dejada por los usuarios al momento de dejar las reseñas. En base a dicha clasificación, se juzga la polaridad de los aspectos identificados. En la vida real, hay casos en donde la clasificación dada por los comensales es positiva pero la polaridad del texto en si es negativa. El usuario pudo haber dejado el comentario con un tono satírico. Se recomienda para futuros trabajos entrenar el modelo con datos que hayan sido manualmente validados para evitar eso.
- El modelo detecta bien los aspectos, sin embargo, para futuros trabajos se considera necesario construir un diccionario único con palabras propias del contexto del restaurante a analizar para mejorar la identificación de aspectos. Este diccionario debería traer por ejemplo nombres de platillos del lugar. Se sugiere esto debido a que los etiquetadores gramaticales usados son generales y en ocasiones no se logran identificar palabras únicas.
- Siguiendo la misma lógica de la recomendación anterior, se sugiere generar un diccionario único para el contexto de los restaurantes para una mejor clasificación del sentimiento. Esto debido a que existen palabras bajo el contexto de los restaurantes pueden tener una polaridad opuesta. Palabras como ‘caliente’, ‘helado’, ‘blando’, etc. son unos ejemplos.

## **CAPITULO VI. APLICABILIDAD**

### **6.1 TÍTULO**

Aplicación del análisis de sentimiento basado en los aspectos culinarios de un restaurante

### **6.2 JUSTIFICACIÓN**

En base al trabajo realizado en esta tesis, cuyos resultados fueron expuestos en el capítulo 4, se afirma que la presencia en línea de los restaurantes les da una oportunidad para la recolección de opiniones que los clientes dejan. Sin embargo, es difícil hacer sentido de dicha información no estructurada. El cliente hace uso de un lenguaje natural para dejar la reseña u opinión lo que lo hace parecer como si se tuviese una entrevista de enfoque individual por cada una. En casos así, potencialmente el negocio trataría con cientos o miles de personas dejando su opinión haciendo uso de su lenguaje natural – obteniendo así el muestreo necesario para un estudio cuantitativo, manteniendo el enfoque profundo de información específica a un estudio cualitativo haciendo uso de un algoritmo de clasificación del sentimiento.

Tradicionalmente, las personas tomaban sugerencias de conocidos, amigos y familiares. En la edad actual, el internet se ha convertido en la principal manera de obtener sugerencias. En estas plataformas en línea, los usuarios pueden compartir sus opiniones y leer la de otras personas acerca el servicio, comida, precio y ambiente de un restaurante. Según el Local Consumer Review Survey 2020 de BrightLocal, el 93% de los clientes de los restaurantes/café ven las reseñas del lugar antes de decidir si ir o no. Este porcentaje es el más alto de cualquier otra industria. La mayoría de los clientes (63%), hace uso de Google para encontrar información acerca de negocios locales, incluyendo reseñas, información de contacto, horarios de apertura, etc. Solo el 48% consideraría un lugar que tenga menos de 4 estrellas. En conclusión, asegurar buenas reseñas de parte de los clientes y poder identificar rápidamente las razones de inconformidad de ellos se ha convertido en un punto clave para asegurar un crecimiento sano del restaurante, traduciéndose a un incremento de ganancias.

### **6.3 ALCANCE DE LA PROPUESTA**

Este trabajo tiene como objetivos:

1. Implementar un análisis del sentimiento y categorización de temas de aspecto culinarios en la retroalimentación cualitativa obtenida de los consumidores del negocio en línea, obteniendo así un análisis cuantitativo que ayude a la toma de decisiones de negocio.
2. Categorizar cada reseña identificada en cuatro aspectos culinarios predefinidos.
3. Identificar mediante el análisis del sentimiento la polaridad de los aspectos o temas definidos en las reseñas de los comensales
4. Determinar las oportunidades de mejora en los aspectos según el análisis del sentimiento de los comentarios dejados por los clientes

### **6.4 DESCRIPCIÓN Y DESARROLLO A DETALLE DE LA PROPUESTA**

#### **6.4.1 DESCRIPCIÓN CLARA DEL QUE Y COMO**

Para el desarrollo de un algoritmo de clasificación del sentimiento y categorización de aspectos culinarios de las reseñas obtenidas en línea de un restaurante se comienza primero con la recolección de datos, las llamadas reseñas. Esta recolección de datos se realizó haciendo uso de la plataforma web Apify, que permite la automatización de este tipo de actividades mediante ‘actores’ desarrollados por la misma plataforma o por terceros. Por medio de Apify, se obtuvieron las reseñas dejadas por los clientes a lo que va de septiembre del 2023 de las diferentes sucursales del restaurante Denny’s en Honduras.

El siguiente paso es la limpieza de los datos mediante el lenguaje Python. Se realiza un preprocesamiento que busca darle una estructura más legible a los datos obtenidos que permita una mejor comprensión y un resultado más certero del algoritmo al momento de realizar la clasificación del sentimiento y la categorización de los aspectos culinarios definidos como: precio, comida, lugar y servicio. En este paso, se eliminan reseñas duplicadas, caracteres no alfabéticos, se normaliza el texto convirtiendo las palabras con acento a su versión sin acento. Realizado esto, se procede a la lematización de las palabras para así poder disminuir la cantidad de información a analizar y a la eliminación de palabras de vacío haciendo uso de la librería Spacy de Python.

Al obtener la data procesada, se procede a la categorización de los aspectos de la reseña haciendo uso de método de LDA (Latent Dirichlet Allocation). La idea básica del LDA es que un documento, en este caso reseñas, consiste en uno o varios aspectos o temas latentes donde cada uno consiste en una distribución de palabras que los identifica. El método LDA trabaja mediante dos suposiciones claves: los documentos son una mezcla de temas y los temas son una mezcla de tokens (palabras). En un lenguaje estadístico, los documentos son conocidos como la distribución probabilística de temas y los temas como la distribución probabilística de palabras. Cada documento es una reseña, y el corpus es la colección de documentos o reseñas alimentados. Mediante la LDA se pueden obtener los temas o aspectos culinarios de cada reseña con un peso dado en cada aspecto definido.

Después de obtener la categorización con su peso individual de cada aspecto por reseña, se procede al entrenamiento del algoritmo de clasificación del sentimiento. En este trabajo se ejecutaron diferentes algoritmos populares de clasificación para poder obtener la polaridad de cada reseña. Se dividió el conjunto de reseñas obtenidas en una relación 70/30, en donde el 70% de las reseñas fueron destinadas al entrenamiento del algoritmo y el 30% se destinaron como prueba.

Al obtener los resultados de cada algoritmo se procedió a realizar su respectiva matriz de confusión para determinar el mejor método. Una vez obtenido, se procedió a identificar las reseñas cuya fecha de publicación se definió desde enero 2022 hasta la fecha actual. En cada reseña, obtuvo la categorización de los aspectos culinarios definidos con su respectivo peso cada uno. En base a ese peso, se calculó la polaridad de cada uno haciendo uso de la clasificación del sentimiento (positivo o negativo) obtenida de cada reseña.

#### 6.4.2 DESARROLLO DE TODOS LOS ELEMENTOS NECESARIOS

Tal y como se definió en la sección previa, en esta sección se lleva a cabo la ejecución de cada paso necesario para la aplicación del análisis de sentimiento basado en los aspectos culinarios de un restaurante

Se concluye que el algoritmo que dio mejores resultados, identificados mediante la ponderación F1 de la matriz de confusión, fue el de regresión logística.

## 6.4 MEDIDAS DE CONTROL

Se hará uso de la matriz de confusión para asegurar que el modelo seleccionado tenga una buena precisión al momento de clasificar el sentimiento de las reseñas.

## 6.5 GUÍA DE IMPLEMENTACIÓN PARA ANALISIS DE SENTIMIENTOS BASADOS EN ASPECTOS EN RESTAURANTES

El análisis de sentimientos basado en aspectos es un trabajo de procesamiento de lenguaje natural que conlleva identificar y analizar el sentimiento expreso hacia diferentes aspectos de un producto o servicio en texto. Latent Dirichlet Allocation (LDA) es una técnica de modelado de tópicos que se puede usar para este propósito, identificando los temas o aspectos en el texto y analizando el sentimiento percibido en esos aspectos. Lo siguiente es una guía con pasos para desarrollar un análisis de sentimiento usando LDA en Python:

### 6.5.1 RECOLECCIÓN DE DATOS

Hay que definir la fuente de las reseñas del restaurante. Hay numerosas fuentes disponibles como ser Facebook, Google Reviews, Instagram y Twitter. Es posible utilizar más de una fuente de datos. Hay que definir el intervalo de tiempo a analizar y comenzar con la extracción de los datos. Hay plataformas como Apify que permiten la extracción de estos datos de manera gratuita con un límite mensual bastante generoso. Si no se desea usar Apify o se sobrepasan los límites de este, se pueden usar librerías de Python especializadas como *Beautiful Soup*.

### 6.5.2 PREPROCESAMIENTO DE DATOS

Se debe preprocesar la data recolectada removiendo las palabras vacías, puntuaciones y caracteres especiales. Se recomienda usar realizar un lematizado para disminuir la cantidad de palabras únicas a analizar. Para realizar estos pasos, hay librerías de Python como *Spacy* y *NLTK* que facilitan el trabajo a realizar. Estas librerías cuentan con modelos preentrenados en diferentes idiomas, incluidos el castellano. En la Ilustración 17, vemos un ejemplo de cómo usar Spacy para remover las palabras vacías haciendo uso de su librería 'es\_core\_news\_md'.

```
import spacy
from unicode import unicode
import re

sp = spacy.load('es_core_news_md')

stop_words = sp.Defaults.stop_words

exclude_stopwords = ['no', 'mejor']
for word in exclude_stopwords:
    stop_words.remove(word)
```

### 6.5.3 APLICAR LDA PARA IDENTIFICAR LOS ASPECTOS

LDA es un modelo probabilístico para el modelado de tópicos. Se puede aplicar para identificar los aspectos en el corpus. Hay librerías como *Gensim* o *Sklearn* que permiten realizar esta tarea. Una vez realizado el modelado, se inspeccionan las palabras claves de los aspectos.

En la Ilustración 18 mostramos el código a utilizar para la clasificación de aspectos. Los parámetros del *CountVectorizer* son los siguientes:

- *Max\_df*: este parámetro se utiliza para remover términos en las reseñas que aparecen demasiado frecuente. Un valor en decimal significa que se ignoraran las palabras que aparezcan en el porcentaje correspondiente del total de reseñas. Un valor entero significa que se ignoraran las palabras que aparezcan más de esa cantidad de veces. En el ejemplo, se usa 0.7, lo que significa que las palabras que aparezcan en más del 70% de las reseñas serán ignoradas.
- *Min\_df*: este parámetro es similar en funcionamiento a *Max\_df*, solamente que este trata con las palabras que aparecen muy poco. En la Ilustración 14, tenemos el número entero 5, lo que significa que las palabras que aparezcan en menos de 5 diferentes reseñas serán ignoradas.

Estos parámetros son importantes para el optimizado del modelado de aspectos y se deben usar libremente si los aspectos identificados por LDA son considerados inferiores.

Ilustración 18 Ejemplo de LDA para la clasificación de aspectos en Python

```
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.decomposition import LatentDirichletAllocation
import random

def get_words(dict, search_index):
    return list(dict.keys())[list(dict.values()).index(search_index)]

reviews_datasets = df_reviews

count_vect = CountVectorizer(max_df=0.7, min_df=5, stop_words=['no'])
doc_term_matrix = count_vect.fit_transform(reviews_datasets['cleaned_reviews'].values.astype('U'))

LDA = LatentDirichletAllocation(n_components=4, random_state=42, learning_decay=0.7)
LDA.fit(doc_term_matrix)

print(LDA.get_params())
i=0
topic_words=[]
for topic in LDA.components_:
    top_topic_words = topic.argsort()[-10:]
    for x in top_topic_words:
        topic_words.append(get_words(count_vect.vocabulary_,x))
    print('Topic #(): {}'.format(i,topic_words))
    topic_words = []
    i+=1
```

#### 6.5.4 ASIGNACIÓN DEL SENTIMIENTO

Para la asignación de la polaridad para cada aspecto se pueden usar algoritmos como regresión logística y máquina de soporte de vectores o técnicas como VADER. En la Ilustración 19, se muestra un ejemplo de cómo entrenar el modelo de regresión logística para la clasificación del sentimiento. Se uso un método de GridSearch que permite encontrar los parámetros ideales para la optimización del modelo.

Ilustración 19 Ejemplo de Regresión Logística para la clasificación del sentimiento

```
param_grid = {'c_vectorizer__ngram_range': [(1, 1), (2, 2), (3, 3), (1, 2), (2, 3), (1, 3)]}

lr_pipeline = Pipeline([
    ('c_vectorizer', CountVectorizer()),
    ('lr', LogisticRegression(random_state=42))
])

grid_search_lr = GridSearchCV(lr_pipeline, cv=2, param_grid=param_grid, scoring='f1')
grid_search_lr.fit(X_train, y_train)

print('Best ngram_range:', grid_search_lr.best_estimator_.get_params()['c_vectorizer__ngram_range'])

predictions = grid_search_lr.predict(X_test)
evaluate(y_test, predictions )
```

## 6.6 CRONOGRAMA DE IMPLEMENTACIÓN Y PRESUPUESTO

### 6.5.1 PRESUPUESTO

En esta sección se detallarán los recursos tecnológicos, físicos y humanos que se estiman necesario asumiendo la creación de un pequeño departamento de tecnología e información dentro del negocio. El software de uso principal en este proyecto fue Python, sin embargo, se puede hacer uso de otros lenguajes estadísticos como R. Para motivos de gráficos, en este trabajo se hizo uso de las mismas librerías graficas de Python sin embargo se considerará la adquisición de herramientas como Power BI para cubrir esta necesidad

Tabla 12 Costo de equipo tecnológico

Equipo Tecnológico			
Concepto	Cantidad	Precio Unitario	Total
Laptops	1	L 37,000.00	L 37,000.00
Impresora	1	L 6,000.00	L 6,000.00
			<b>L 43,000.00</b>

Tabla 13 Costo de software

Software			
Concepto	Cantidad	Precio Unitario	Total
Office 365 Empresa Estándar Anual	1	L 3,900.00	L 3,900.00
Power Bi Pro	1	L 3,000.00	L 3,000.00
			<b>L 6,900.00</b>

Tabla 14 Costo de mobiliario

Mobiliario			
Concepto	Cantidad	Precio Unitario	Total
Sillas ejecutivas	1	L 3,000.00	L 3,000.00
Escritorios	1	L 6,000.00	L 6,000.00
			<b>L 27,000.00</b>

Tabla 15 Costo del recurso humano

Recurso Humano		
Concepto	Cantidad	Sueldo Mensual
Analista de información	1	L 24,000.00

### 6.5.2 CRONOGRAMA

A continuación, en la Tabla 16 se detallan las actividades necesarias para complementar la implementación de un modelo de análisis del sentimiento basado en aspectos.

Tabla 16 Cronograma de actividades

No.	Actividad	Semana	1	2	3	4	5	6	7	8	9	10	11	12
1	Contratación de analista		■	■	■	■								
2	Entrenamiento del analista						■	■	■					
3	Determinar la fuente de datos									■				
4	Procesamiento de datos										■			
5	Entrenamiento del modelo										■			
6	Análisis de resultados											■	■	
7	Validación del modelo											■	■	
8	Implementación del modelo													■

En el cronograma de implementación se detallan las acciones a realizar de acuerdo con las semanas de ejecución. Se define la necesidad de contratación de un analista con un perfil capaz de desarrollar un modelo de análisis de clasificación del sentimiento usando lenguajes como Python o R. Se requiere un entrenamiento de este para familiarizarse con su entorno y poder obtener los requisitos según las necesidades del negocio.

Después de la contratación y entrenamiento del analista, se procede a determinar la fuente de datos a utilizar. En este caso se busca obtener la retroalimentación de los clientes del restaurante dejadas en sus redes sociales.

Con la fuente de datos obtenida, se procede a la limpieza de esta. Aplicando las técnicas estándares de limpieza de datos para asegurar así un buen entrenamiento del modelo a realizar. Después se procede con la creación del modelo en sí y su consecuente entrenamiento.

Al obtener los primeros resultados del modelo, es necesario realizar un análisis exhaustivo para asegurar la veracidad de este. Se procede también con la validación del modelo aplicando los instrumentos de medición establecidos. Ya validado el modelo, se procede finalmente con la implementación de este al obtener resultados satisfactorios.

## REFERENCIAS

- Ahmed, S., Asheq, A. A., Ahmed, E., Chowdhury, T. S., & Mostafa, G. (2022). The intricate relationships of consumers' loyalty and their perceptions of service quality, price and satisfaction in restaurant service. *The TQM Journal*, 519-539.
- Baquedano, K. (2022, febrero 18). *El 48% de las empresas de Honduras vende por Internet*. Retrieved from LaPrensa: <https://www.laprensa.hn/honduras/el-48-de-las-empresas-de-honduras-vende-por-internet-CH5876334>
- Barrios, J. (2019, Julio 26). *La matriz de confusión y sus métricas*. Retrieved from JuanBarrios.com: <https://www.juanbarrios.com/la-matriz-de-confusion-y-sus-metricas/>
- Bezco, G. (2023, Junio 16). *Lexicon-based vs. Machine Learning-based Sentiment Analysis For Contact Centers*. Retrieved from miarec: <https://blog.miarec.com/lexicon-based-vs-ml-based-sentiment-analysis>
- BID. (2021, Junio). *Brecha Digital en America Latina y el Caribe*. Retrieved from Banco Interamericano de Desarrollo: <https://publications.iadb.org/es/informe-anual-del-indice-de-desarrollo-de-la-banda-ancha-idba-2020-brecha-digital-en-america-latina>
- Čurn, J. (2019, Abril 9). *The Apify Master Plan*. Retrieved from blog.apify.com: <https://blog.apify.com/the-apify-master-plan/>
- Diario La Prensa. (2023, mayo 21). *Todo Honduras tendrá Internet satelital de Starlink en 2024*. Retrieved from laprensa.hn: <https://www.laprensa.hn/honduras/honduras-tendra-internet-satelital-starlink-2024-GD13545298>
- Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, Vol 69 (No. 2), 897-904. Retrieved from <https://www.sciencedirect.com/science/article/abs/pii/S0148296315002842>
- Goncalves, P., Araujo, M., Benevenuto, F., & Cha, M. (2013). Comparing and Combining Sentiment Analysis Methods. *ACM Conference on Online Social Networks*. Boston. Retrieved Febrero 10, 2023, from <https://arxiv.org/pdf/1406.0032.pdf>
- IBM. (2021, Agosto 17). *IBM*. Retrieved from Funcionamiento de SVM: <https://www.ibm.com/docs/es/spss-modeler/saas?topic=models-how-svm-works>
- IBM. (2023, Agosto 4). *IBM*. Retrieved Septiembre 14, 2023, from Regresión Logística: <https://www.ibm.com/docs/es/spss-statistics/saas?topic=regression-logistic>
- Mahesh, J., Dipanjan, D., Kevin, G., & Noah A., S. (2010). *Movie Reviews and Revenues: An Experiment in Text Regression*. Retrieved from aclanthology.org: <https://aclanthology.org/N10-1038.pdf>
- Mesa Editorial. (2022, Diciembre 9). *Honduras tiene mas de 8 millones de usuarios de telefonía móvil, 6.6% mas que en 2021*. Retrieved from Dinero: <https://dinero.hn/honduras-tiene-mas-de-8-millones-de-usuarios-de-telefonía-móvil-66-mas-que-en-2021/>
- Mo, Y., Kontonatsios, G., & Ananiadou, S. (2015). Supporting systematic reviews using LDA-based document representations. *Systematic Reviews*, 172. doi:<https://doi.org/10.1186/s13643-015-0117-0>

- Moreno, F., Martelo, C., Corredor, B., Cifuentes, J., & Lopez, J. P. (2020). Técnicas para la Clasificación de Sentimientos en Redes Sociales como Apoyo en el Marketing Digital. *Revista Ibérica de Sistemas e Tecnologías de Información*, 167-186.
- Nurifan, F., Sarno, R., & Sungkono, K. R. (2019). Aspect Based Sentiment Analysis for Restaurant Reviews Using Hybrid ELMo-Wikipedia and Hybrid Expanded Opinion Lexicon-SentiCircle. *International Journal of Intelligent Engineering & Systems*, 47-58.
- Observatorio Demográfico Universitario. (2021). *Boletín 11: Centroamérica “Tendencias de las Tecnologías de la Información y Comunicación 2010-2019*. Retrieved from ODU UNAH: <https://odu.unah.edu.hn/publicaciones/boletines/>
- Omar Andres, C. G. (2019, Diciembre 13). *Desarrollo de una aplicación web para la identificación de oportunidades de emprendimiento e inversión utilizando análisis de sentimientos de Tweets*. Retrieved from Unipiloto: <http://repository.unipiloto.edu.co/handle/20.500.12277/7778>
- Ortega, C. (n.d.). *Investigación cuantitativa. Qué es y cómo realizarla*. Retrieved Marzo 12, 2023, from QuestionPro: <https://www.questionpro.com/blog/es/que-es-la-investigacion-cuantitativa/>
- Pérez Porto, J., & Merino, M. (2021, Noviembre 10). *Qué es, definición y concepto*. Retrieved from Definicion.de: <https://definicion.de/comida/>
- Peter, T. (2002, Julio). *Thumbs Up or Thumbs Down? Semantic Orientation Applied to*. Retrieved from aclanthology.org: <https://aclanthology.org/P02-1053.pdf>
- Samet, A. (2020, Junio 9). *2020 US SOCIAL MEDIA USAGE: How the Coronavirus is Changing Consumer Behavior*. Retrieved from Business Insider: <https://www.businessinsider.com/2020-us-social-media-usage-report>
- Scherer, K. R. (1984). Emotion as a multicomponent process: A model and some cross-cultural data. *Review of Personality & Social Psychology*(5), 37-63.
- Secretaria Educacion. (2021, Enero 13). *Gobierno brindará Internet gratuito a unos 75,000 hogares hondureños*. Retrieved from Secretaria de Educacion HN: <https://www.se.gob.hn/detalle-articulo/1554/>
- Srivastava, R., Bharti, P. K., & Verma, P. (2022). Comparative Analysis of Lexicon and Machine Learning Approach for Sentiment Analysis. *International Journal of Advanced Computer Science and Applications*, 13(3). Retrieved from [https://thesai.org/Downloads/Volume13No3/Paper\\_12-Comparative\\_Analysis\\_of\\_Lexicon\\_and\\_Machine\\_Learning\\_Approach.pdf](https://thesai.org/Downloads/Volume13No3/Paper_12-Comparative_Analysis_of_Lexicon_and_Machine_Learning_Approach.pdf)
- Starlink. (n.d.). *Starlink*. Retrieved Octubre 31, 2023, from starlink: <https://www.starlink.com/>
- Statista. (2021, Octubre 22). *El Big Bang del Big Data*. Retrieved from Statista: <https://es.statista.com/grafico/26031/volumen-estimado-de-datos-digitales-creados-o-replicados-en-todo-el-mundo/>
- Thota, S. C. (2018). *Social Media: A Conceptual Model of the Whys, Whens and Hows of Consumer Usage of Social Media and Implications on Business Strategies*. Retrieved from Academy of Marketing Studies Journal : <https://www.abacademies.org/articles/social-media-a-conceptual-model-of-the-whys-whens-and-hows-of-consumer-usage-of-social-media-and-implications-on-business-strategi-7531.html>

Tumasjan, A., Sprenger, T., Sandner, P., & Weppe, I. (2010, Mayo 16). *Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment*. Retrieved from ojs.aaai.org:  
<https://ojs.aaai.org/index.php/ICWSM/article/view/14009>

## ANEXOS

### ANEXO 1: CONJUNTO DE DATOS

publishedAtDate	reviews	sentiment
2023-09-08	Denny's es un lugar que frecuentó a menudo porque es único en su categoría. Su	1
2023-09-07	Precios cómodos Buena atención	1
2023-09-06	No pidan los samples. Los boneless vienen crudos. Entradas que arruinan el ape	0
2023-09-04	Hoy estuvo muy bien	1
2023-09-03	Pésima atención. Mas de 1hora esperando para un desayuno para dos personas	0
2023-09-01	Excelente lugar para ir a desayunar	1
2023-09-01	Muy buen servicio, y excelente ambiente	1
2023-09-01	Pedí la comida primero y me dejaron de último nisiquiera encontraban mi pedid	1

### ANEXO 2: MATRIZ DE CONFUSIÓN

		Polaridad Real	
		Postividad	Negatividad
Polaridad Predicha	Postividad	Verdadero Positivo	Falso Positivo
	Negatividad	Falso Negativo	Verdadero Negativo

### ANEXO 3: RESULTADOS DE METRICA

Medicion %	Aprendizaje Automatico
Exactitud	
Precision	
Sensibilidad	
Puntaje F1	