

**FACULTAD DE ESTUDIOS ESTADÍSTICOS**

**MÁSTER EN CIENCIA DE DATOS E INTELIGENCIA  
DE NEGOCIOS**

Curso 2024/2025

---

**Trabajo de Fin de Máster**

***TÍTULO:* Análisis de Sentimiento Basado en Aspectos (ABSA) aplicado a reseñas de restaurantes: Un estudio con modelos Transformer en Dianping**

***Alumno:* Julián Darío Reina Correa**

***Tutores:* Belén Rodríguez Cánovas**

Julio de 2025



UNIVERSIDAD COMPLUTENSE  
MADRID

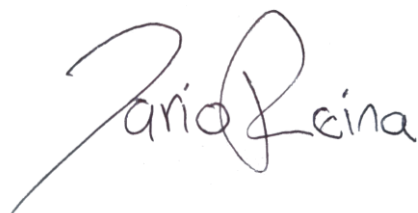
## Declaración Responsable sobre Autoría y Uso Ético de Herramientas de Inteligencia Artificial (IA)

Yo, **Julián Darío Reina Correa** con DNI/NIE/PASAPORTE: **Z2217145L** declaro de manera responsable que el/la presente **Trabajo de Fin de Máster (TFM)** titulado/a: **Análisis de Sentimiento Basado en Aspectos (ABSA) aplicado a reseñas de restaurantes: Un estudio con modelos Transformer en Dianping**, es el resultado de mi trabajo intelectual personal y creativo, y ha sido elaborado de acuerdo con los principios éticos y las normas de integridad vigentes en la comunidad académica y, más específicamente, en la Universidad Complutense de Madrid.

Soy, pues, autor del material aquí incluido y, cuando no ha sido así y he tomado el material de otra fuente, lo he citado o bien he declarado su procedencia de forma clara -incluidas, en su caso, herramientas de inteligencia artificial-. Las ideas y aportaciones principales incluidas en este trabajo, y que acreditan la adquisición de competencias, son mías y no proceden de otras fuentes o han sido reescritas usando material de otras fuentes.

Asimismo, aseguro que los datos y recursos utilizados son legítimos, verificables y han sido obtenidos de fuentes confiables y autorizadas. Además, he tomado medidas para garantizar la confidencialidad y privacidad de los datos utilizados, evitando cualquier tipo de sesgo o discriminación injusta en el tratamiento de la información.

En Madrid a 01 de julio de 2025



---

## Resumen

El presente trabajo explora la aplicación de técnicas avanzadas de análisis de sentimiento basado en aspectos (ABSA) sobre reseñas textuales de restaurantes, con especial foco en la plataforma Dianping. Para ello, se utilizó un pipeline que integra modelos de lenguaje basados en Transformer (BERT y DeBERTa), permitiendo analizar tanto la polaridad global como la percepción detallada en aspectos clave como sabor, servicio y ambiente.

Los resultados revelan una tendencia general positiva en la mayoría de las reseñas, con más del 90% de opiniones favorables detectadas automáticamente. Sin embargo, la comparación con las puntuaciones explícitas mostró discrepancias importantes, especialmente en las categorías neutra y negativa. Además, el análisis ABSA permitió identificar correlaciones significativas entre los *scores* numéricos y los sentimientos inferidos, destacando el servicio como el aspecto con mayor coherencia.

El estudio demuestra el valor añadido de ABSA y modelos Transformer para captar matices que no siempre son evidentes en las valoraciones numéricas, ofreciendo una herramienta poderosa para comprender mejor la satisfacción del cliente en el sector gastronómico.

**Palabras clave:** NLP, Análisis de sentimiento, ABSA, Transformer, reseñas online, Dianping, satisfacción del cliente.

---

## Abstract

This study explores the application of advanced aspect-based sentiment analysis (ABSA) techniques to restaurant reviews, focusing on the Dianping platform. A pipeline integrating Transformer-based language models (BERT and DeBERTa) was used to analyze both overall polarity and detailed perception across key aspects such as taste, service, and environment.

The results reveal a general positive tendency in most reviews, with over 90% of favorable opinions automatically detected. However, comparison with explicit numerical scores showed important discrepancies, especially in neutral and negative categories. Additionally, ABSA analysis identified significant correlations between numerical scores and inferred sentiments, with service emerging as the most consistent aspect.

This work demonstrates the added value of ABSA and Transformer models in capturing subtle nuances not always reflected in numerical ratings, providing a powerful tool to better understand customer satisfaction in the restaurant sector.

**Keywords:** NLP, Sentiment analysis, ABSA, Transformer, online reviews, Dianping, customer satisfaction.

---

# Índice

<b>1. Introducción</b>	<b>6</b>
<b>2. Objetivos de Investigación</b>	<b>9</b>
2.1. Objetivo general . . . . .	9
2.2. Objetivos específicos . . . . .	9
<b>3. Marco Teórico</b>	<b>10</b>
3.1. Análisis de Sentimiento Basado en Aspectos (ABSA) . . . . .	10
3.2. Enfoques clásicos: reglas lingüísticas, diccionarios y modelos de aprendizaje automático . . . . .	13
3.3. Modelos de Deep Learning . . . . .	16
3.3.1. Redes neuronales tradicionales . . . . .	16
3.3.2. Transformers, BERT y DeBERTa . . . . .	19
3.4. Particularidades del análisis de texto en chino . . . . .	22
<b>4. Revisión de literatura</b>	<b>23</b>
4.1. Revisión de literatura industria . . . . .	23
4.2. Revisión de literatura Dianping.com . . . . .	25
<b>5. Metodología</b>	<b>29</b>
5.1. Descripción del conjunto de datos . . . . .	29

---

5.2.	Limpieza y preprocesamiento de datos . . . . .	31
5.3.	Tokenización y análisis léxico preliminar . . . . .	32
5.4.	Etiquetado de sentimientos por <i>Rating</i> . . . . .	33
5.5.	Ajuste fino del modelo BERT . . . . .	33
5.5.1.	Configuración y entrenamiento . . . . .	34
5.6.	Evaluación y aplicación del modelo . . . . .	35
5.6.1.	Comparación entre etiquetas por <i>rating</i> y predicción BERT . . . . .	35
5.7.	Análisis ABSA . . . . .	36
5.7.1.	Extracción y agrupación de aspectos . . . . .	36
5.7.2.	Aplicación del modelo DeBERTa . . . . .	37
5.7.3.	Análisis cuantitativo y estadístico por aspecto . . . . .	37
5.8.	Resumen del <i>pipeline</i> metodológico . . . . .	38
<b>6.</b>	<b>Resultados</b>	<b>40</b>
6.1.	Análisis descriptivo del conjunto de datos . . . . .	40
6.2.	Análisis léxico preliminar . . . . .	43
6.3.	Etiquetado y análisis de sentimiento global . . . . .	45
6.4.	Evaluación del modelo de sentimiento . . . . .	46
6.5.	Comparación entre etiquetas de <i>Rating</i> y predicción BERT . . . . .	49
6.6.	Análisis de aspectos y resultados ABSA . . . . .	50

---

6.7. Análisis cuantitativo y estadístico . . . . .	54
<b>7. Discusión</b>	<b>58</b>
<b>8. Limitaciones</b>	<b>59</b>
<b>9. Conclusiones</b>	<b>61</b>
<b>10. Trabajos futuros</b>	<b>62</b>
<b>11. Bibliografía</b>	<b>64</b>

---

# 1. Introducción

El Procesamiento del Lenguaje Natural (Natural Language Processing, NLP) se ha consolidado como uno de los campos más demandados de la inteligencia artificial dado que su objetivo principal es permitir a las máquinas comprender, interpretar y generar lenguaje humano, lo que facilita la interacción entre personas y sistemas computacionales (Jurafsky y Martin 2021). Los modelos de NLP pueden ser usados en múltiples escenarios, como las traducciones automáticas, reconocimiento de entidades nombradas, generación de texto, clasificación de documentos y análisis de sentimientos (Bird, Klein y Loper 2009). Una de sus principales fortalezas es la capacidad para extraer información relevante de data no estructurada como pueden ser las reseñas, publicaciones en redes sociales o cualquier otro tipo de contenido generado por los usuarios (Cambria et al. 2017).

Para mejorar los resultados en NLP, se han desarrollado arquitecturas más sofisticadas, capaces de comprender el significado tanto a nivel de palabra como de frase, y de identificar fenómenos complejos como la ironía, la ambigüedad o la subjetividad (Young et al. 2018). Este tipo de análisis resulta especialmente relevante para sectores como el comercio electrónico, la hostelería y los servicios, donde las reseñas representan el principal mecanismo de retroalimentación y ajuste de oferta. (Medhat, Hassan y Korashy 2014).

Estas reseñas pueden obtenerse de plataformas especializadas en la evaluación y recomendación de establecimientos comerciales, como Yelp, TripAdvisor o Dianping. En este trabajo se analiza un gran volumen de reseñas textuales extraídas de Dianping.com, la principal plataforma de reseñas de restaurantes y servicios en China. Fundada en 2003, Dianping está ampliamente reconocida como referente para recomendaciones de restaurantes, hoteles y otros establecimientos. Los usuarios pueden valorar aspectos como el servicio, la comida y el ambiente, y el sistema calcula una puntuación global basada en estos criterios mediante una escala de cinco estrellas (dianping.com 2022).

En este contexto, una de las aplicaciones con mayor relevancia tanto académica como empresarial es el análisis de sentimientos (sentiment analysis, SA) también conocido como mi-

---

nería de opiniones (Cambria et al. 2017). De acuerdo con Liu (2012) el SA suele emplearse para identificar la orientación emocional de un texto y asignarle una polaridad, evaluándose entre positivo, neutro o negativo. El estudio del SA es posible a tres niveles, dependiendo del nivel de detalle requerido.

1. Nivel de documento: Evalúa la polaridad general de un texto completo.
2. Nivel de oración: Determina el sentimiento en frases individuales dentro de un documento.
3. Nivel de aspecto: Analiza opiniones sobre características específicas de un producto o servicio.

Dada la naturaleza de las reseñas y el nivel de granularidad necesario para evaluarlas, las investigaciones recientes se han enfocado en el análisis de sentimiento basado en aspectos (Aspect-Based Sentiment Analysis, ABSA) que resuelve este problema al descomponer la evaluación en distintos aspectos del objeto evaluado, de modo que, por ejemplo, en una reseña de restaurante el precio puede recibir una evaluación positiva mientras que el servicio sea percibido negativamente. Este nivel de detalle proporciona información más detallada y útil tanto para consumidores como para empresas (Pontiki et al. 2014).

Un avance significativo en NLP llegó con la introducción del modelo Transformer por Vaswani et al. (2017) en el artículo "Attention Is All You Need". Este modelo reemplaza la recurrencia con un mecanismo de autoatención (self-attention), permitiendo procesar palabras en paralelo y capturar relaciones contextuales de manera más eficiente. Gracias al uso de arquitecturas tipo Transformer, el campo del ABSA ha experimentado una mejora radical. Modelos como BERT han demostrado una capacidad sobresaliente para capturar relaciones contextuales, lo que mejora notablemente el rendimiento en la comprensión semántica (Xu et al. 2020).

Por lo tanto, este trabajo presenta una metodología reproducible y específicamente adaptada al idioma chino para el ABSA multiclase en reseñas de restaurantes de Dianping.com, empleando modelos Transformer de última generación. A diferencia de investigaciones previas,

---

la propuesta abarca desde el preprocesamiento detallado de un corpus real, hasta la extracción semi-automatizada y agrupación de aspectos frecuentes. Además, se integra un análisis cuantitativo exhaustivo que permite relacionar las valoraciones explícitas de los usuarios con los sentimientos detectados automáticamente en el texto. Los resultados obtenidos contribuyen al conocimiento académico sobre ABSA y proporcionan herramientas y recomendaciones prácticas para su aplicación en entornos reales.

El presente trabajo se estructura de la siguiente manera: en la Sección 2 se presentan los objetivos de la investigación; en la Sección 3 se expone el marco teórico; en la Sección 4 se desarrolla la revisión de la literatura; en la Sección 5 se describe detalladamente la metodología empleada; en la Sección 6 se exponen y analizan los resultados obtenidos; en la Sección 7 se desarrolla la discusión de los hallazgos; en la Sección 8 se abordan las principales limitaciones del estudio; en la Sección 9 se presentan las conclusiones; y finalmente, en la Sección 10 se proponen líneas para trabajos futuros.

---

## 2. Objetivos de Investigación

### 2.1. Objetivo general

Realizar un ABSA en reseñas de restaurantes de Dianping, aplicando modelos de *deep learning* y técnicas modernas de procesamiento de texto en chino, para lograr una interpretación detallada y visual de la experiencia del usuario.

### 2.2. Objetivos específicos

- Analizar el sentimiento general de las reseñas de restaurantes de Dianping utilizando modelos de *deep learning* adaptados al procesamiento de texto en chino.
- Clasificar el sentimiento asociado a los aspectos relevantes presentes en las reseñas mediante técnicas de ABSA.
- Explorar la relación cuantitativa entre los sentimientos por aspecto y las valoraciones tradicionales.

---

### 3. Marco Teórico

En esta sección se abordan los conceptos teóricos y académicos fundamentales en los que se enmarca la realización de este trabajo. En la Sección 3.1 se habla del ABSA en el marco del NLP, abordando su papel como nivel de análisis de sentimientos, aplicaciones industriales y ejemplos relevantes. En la Sección 3.2 se habla de los modelos clásicos y las tecnologías usadas históricamente para evaluar sentimientos en texto. En la Sección 3.3 se explican los principios de la arquitectura Transformer, el modelo BERT y su evolución DeBERTa, resaltando sus ventajas frente a enfoques anteriores. Finalmente, la Sección 3.4 describe las particularidades del análisis de texto en chino y la adaptación de estas tecnologías a dicho contexto.

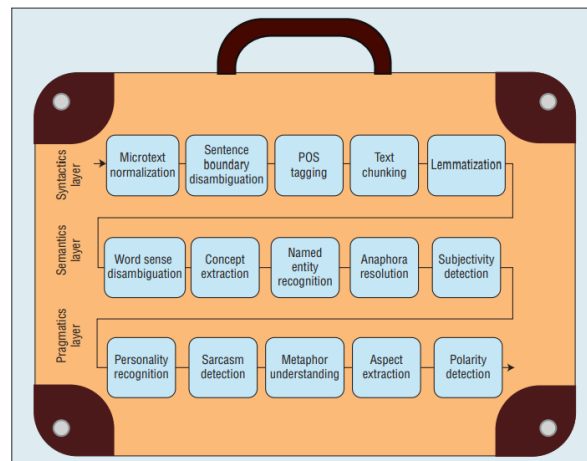
#### 3.1. Análisis de Sentimiento Basado en Aspectos (ABSA)

En el panorama de la Inteligencia Artificial, el Procesamiento de Lenguaje Natural (NLP) surge como una disciplina que se ocupa de la interacción entre los ordenadores y el lenguaje humano (Jurafsky y Martin 2021). Su objetivo es permitir que las máquinas comprendan, interpreten y generen textos escritos o hablados de forma que resulten útiles para tareas concretas, como la traducción automática, el análisis de sentimientos, la extracción de información y la generación automática de texto, entre otras (Bird, Klein y Loper 2009).

En el contexto de este trabajo, el NLP proporciona las herramientas fundamentales para analizar grandes volúmenes de reseñas textuales, que pueden ser de utilidad en diversas aplicaciones como en el SA (Jurafsky y Martin 2021). El SA, también conocido como *opinion mining*, tiene como objetivo identificar, extraer y analizar automáticamente la información subjetiva expresada en los textos. Esta tarea permite clasificar opiniones, actitudes o emociones, facilitando la interpretación de grandes volúmenes de datos generados por los usuarios en distintos contextos (Liu 2012).

A diferencia de otras aplicaciones de NLP, el SA enfrenta retos particulares (véase Figura 1)

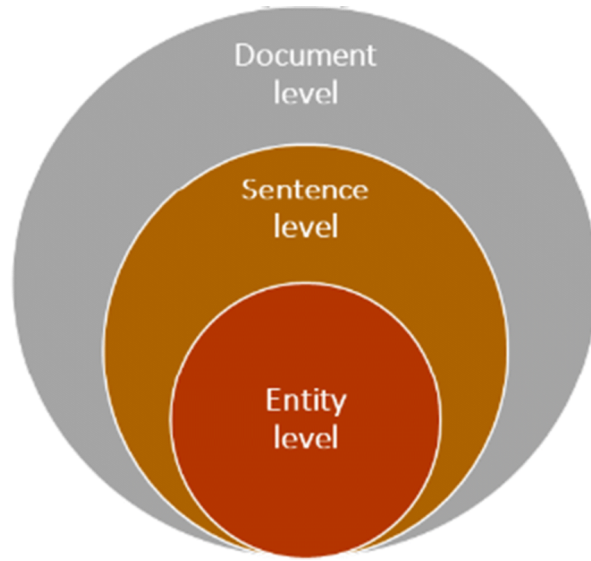
que dificultan alcanzar un desempeño equiparable al humano. Esto se debe a que el SA debe lidiar con fenómenos como la ambigüedad, el sarcasmo, las metáforas y otros aspectos complejos del lenguaje natural, que requieren no solo innovación computacional sino también una comprensión lingüística profunda (Cambria 2017). Sin embargo, su correcta explotación exige un análisis de alta granularidad que permita captar matices específicos (Liu 2012).



**Figura 1:** Principales obstáculos del NLP organizados por niveles. Fuente: Cambria 2017

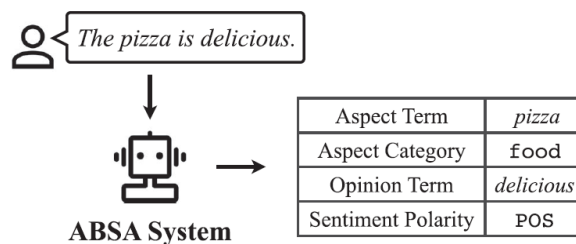
El SA ha sido estudiado en diferentes niveles debido al grado de granularidad necesario en el análisis, como se muestra en la Figura 2. Como esclarece Liu (2012) el nivel documento se caracteriza por considerar que la totalidad de un texto representa un sentimiento negativo o positivo, el nivel de frase analiza el sentimiento de una oración y lo categoriza en positivo, neutro o negativo; finalmente, en el nivel aspecto se busca solucionar la falta de granularidad y, por lo tanto, de precisión de los dos niveles anteriores. En este caso, se entiende que en una opinión está compuesta por el sentimiento y el objeto de dicho sentimiento, por lo que cada aspecto puede tener distinta polaridad. Por ello, para el estudio detallado de reseñas, resulta especialmente recomendable el uso del enfoque ABSA.

Como expone (Zhang et al. 2022), el ABSA se compone de cuatro elementos clave: la categoría del aspecto, el término de aspecto, la expresión de opinión y la polaridad. La categoría del aspecto corresponde a un atributo específico de una entidad dentro de un conjunto predefinido para cada dominio de interés. El término del aspecto se refiere al objetivo de la opinión que aparece explícitamente en el texto. Por su parte, el término de la opinión es la



**Figura 2:** Niveles de Análisis de Sentimientos. Fuente: Pozzi et al. 2016

expresión concreta utilizada por el emisor para manifestar su sentimiento respecto al objetivo, por ejemplo, “deliciosa” en la frase mostrada en la Figura 3. Finalmente, la polaridad del sentimiento describe la orientación de la opinión sobre una categoría o término de aspecto, y suele clasificarse como positiva o negativa si es binaria o positiva, neutra y negativa si es multiclase. Cabe aclarar que algunos autores como Do et al. (2023) consideran las dos primeras una sola tarea, por lo que en este caso serían 3.



**Figura 3:** 4 elementos clave del ABSA. Fuente: Zhang et al. 2022

En la práctica, el enfoque ABSA ha demostrado ser especialmente relevante en sectores donde la experiencia del usuario depende de múltiples dimensiones evaluadas de manera independiente, como ocurre en la industria de la restauración, el turismo o el comercio electrónico (Li et al. 2023). Esta granularidad permite a empresas y organizaciones identificar fortalezas y debilidades concretas a partir de las opiniones de los usuarios y orientar estrategias de mejora basadas en datos objetivos.

---

### 3.2. Enfoques clásicos: reglas lingüísticas, diccionarios y modelos de aprendizaje automático

Como señala Medhat, Hassan y Korashy (2014) y Liu (2012), el SA y por extensión el ABSA ha experimentado una evolución significativa en cuanto a metodologías y técnicas empleadas para su implementación. Los modelos clásicos se fundamentaban en reglas lingüísticas, diccionarios y técnicas tradicionales de aprendizaje automático. A continuación se expondrán los enfoques más relevantes previos a la implementación de modelos de *deep learning*, con sus principales características, usos y limitaciones, dando contexto a los primeros esfuerzos en esta disciplina.

Uno de los enfoques más predominantes en los inicios del SA fueron las reglas lingüísticas, basadas en patrones gramaticales, expresiones regulares y heurísticas diseñadas por lingüistas expertos para identificar la polaridad de un texto. Como se muestra en la Tabla 1, este enfoque era muy limitado y requería un amplio conocimiento lingüístico para su implementación satisfactoria. Además, era poco escalable, ya que dependía del idioma y sus variantes, lo que dificultaba su aplicación estándar y provocaba fallos en contextos complejos sin intervención experta constante (Turney 2002; Hu y Liu 2004).

**Tabla 1:** Ejemplos de reglas lingüísticas para análisis de sentimientos.

Regla	Descripción	Ejemplo de frase	Polaridad
“no” + adjetivo positivo	Negación invierte la polaridad positiva	“No bueno”	Negativo
“muy” + adjetivo negativo	Intensificador aumenta la negatividad	“Muy malo”	Negativo fuerte
[sujeto] + “odia” + [objeto]	Verbos explícitos de opinión negativa	“El cliente odia el servicio”	Negativo
[sujeto] + “recomienda” + [objeto]	Verbos explícitos de opinión positiva	“Recomiendo este restaurante”	Positivo

*Tabla adaptada de Turney 2002 y Hu y Liu 2004.*

Posteriormente surgieron los métodos basados en diccionarios léxicos, que consisten en listas predefinidas por palabras y su respectiva polaridad (positiva, neutra o negativa). Entre los ejemplos más conocidos se encuentran SentiWordNet, el léxico de Hu y Liu y AFINN. Estos enfoques aplican reglas heurísticas sencillas, como sumar las puntuaciones de todas las palabras de un texto o invertir la polaridad tras una negación, para estimar el sentimiento general de la reseña (Hu y Liu 2004). Como se ilustra en la Tabla 2, su principal fortaleza radica en su facilidad de implementación y en que, al ser un método no supervisado, no requiere datos etiquetados. Sin embargo, presentan limitaciones notorias: la falta de sensibilidad al contexto, dificultades para manejar sarcasmo o ambigüedad y una dependencia crítica de la calidad del diccionario utilizado (Hu y Liu 2004).

Paralelamente, surgieron enfoques estadísticos simples basados en la frecuencia de términos positivos y negativos, los cuales tampoco capturan relaciones complejas entre palabras. Algunos trabajos exploraron métodos híbridos que combinan reglas, léxicos y algoritmos sencillos de clasificación, con el objetivo de mejorar la precisión en dominios concretos (Medhat, Hassan y Korashy 2014).

**Tabla 2:** Enfoques de reglas y diccionarios léxicos.

Enfoque	Ejemplo	Características
Basado en reglas	Reglas de patrones gramaticales	Altamente dependiente de expertos, difícil de escalar
Léxico-based	SentiWordNet, Hu y Liu Lexicon, AFINN	No captura contexto, fácil de implementar
Estadístico simple	Conteo de palabras positivas/negativas	Conteo de palabras positivas/negativas
Híbrido	Diccionario + SVM, reglas + NB	Integra reglas/léxicos y ML, mejora precisión en dominios simples

*Tabla adaptada de Hu y Liu 2004.*

El siguiente gran avance, como se muestra en la Tabla 3, fue la adopción de técnicas tradicionales de aprendizaje automático supervisado, como *Naive Bayes*, *Support Vector Machines* (SVM), árboles de decisión, regresión logística y *K-Nearest Neighbors* (KNN) (Pang, Lee y Vaithyanathan 2002; Medhat, Hassan y Korashy 2014). Estos modelos requieren datos etiquetados y transforman el texto en vectores de características mediante métodos como *bag-of-words*, *n-gramas* o *TF-IDF*. Su ventaja reside en su mayor adaptabilidad y capacidad para capturar patrones más complejos. No obstante, aunque representaron una mejora respecto a los métodos léxicos y de reglas, siguen presentando limitaciones importantes, en especial para capturar relaciones semánticas profundas, ya que dependen de una ingeniería de características laboriosa y presentan dificultades ante fenómenos complejos como la ironía o el sarcasmo.

**Tabla 3:** Modelos clásicos de ML

Técnica tradicional	Descripción breve	Ecuación básica
Naive Bayes	Clasificador probabilístico basado en la independencia de las palabras	$P(c d) = \frac{P(c) \prod_{i=1}^n P(w_i c)}{P(d)}$
SVM	Busca el mejor hiperplano para separar clases	$f(x) = w^T x + b$
Árboles de decisión / RF	Basado en reglas jerárquicas, conjunto de árboles	-
Regresión logística	Clasificación binaria/multiclase lineal	$P(y = 1 x) = \frac{1}{1 + e^{-(w^T x + b)}}$
KNN	Clasifica según los “vecinos” más cercanos	$d(x, x_i) = \sqrt{\sum_j (x_j - x_{ij})^2}$
Reglas de asociación	Descubre patrones frecuentes	$A \rightarrow B : \text{support, confidence}$
K-means	Agrupar datos similares no supervisado	$\arg \min_S \sum_{i=1}^k \sum_{x \in S_i} \ x - \mu_i\ ^2$

Tabla adaptada de Pang, Lee y Vaithyanathan 2002 y Medhat, Hassan y Korashy 2014.

---

La literatura también describe enfoques híbridos, combinando diccionarios y modelos supervisados, que permitieron entregar resultados más precisos y proporcionar análisis más útiles en el área (Medhat, Hassan y Korashy 2014; Liu 2012). En conjunto, estos enfoques sentaron las bases para el desarrollo posterior del campo. Sin embargo, sus limitaciones, especialmente para captar matices semánticos y dependencias contextuales, propiciaron la evolución hacia modelos más sofisticados, como los de aprendizaje profundo.

### **3.3. Modelos de Deep Learning**

Los modelos de aprendizaje profundo (*deep learning*, DL) han supuesto una revolución en el análisis de sentimientos y, especialmente, en el ABSA. Frente a las reglas lingüísticas, los diccionarios y los modelos clásicos de aprendizaje automático, las redes neuronales profundas han demostrado una capacidad superior para aprender representaciones ricas del texto, modelar relaciones complejas y adaptarse a contextos variables (Do et al. 2023). En esta sección se revisan las arquitecturas fundamentales del *deep learning* aplicadas al NLP previas a la aparición de los modelos *Transformer*, evidenciando tanto sus aportes como las limitaciones que motivaron la evolución hacia enfoques más avanzados. Finalmente, se aborda la aparición de la arquitectura *Transformer* y algunos modelos derivados relevantes para este trabajo.

#### **3.3.1. Redes neuronales tradicionales**

Antes del auge de los modelos *Transformer*, diversas arquitecturas neuronales permitieron superar varias limitaciones de los enfoques estadísticos y basados en reglas (Do et al. 2023). Como se puede ver en la Tabla 4, a continuación se revisan las redes neuronales tradicionales más relevantes y su papel en el avance del campo.

En este panorama, los primeros modelos implementados en análisis de sentimientos fueron las redes neuronales artificiales (*ANN*, *DNN* o *MLP*), compuestas por varias capas densas. Aunque permitieron experimentar con arquitecturas profundas, su capacidad para capturar

---

relaciones contextuales o dependencias secuenciales era limitada, ya que procesaban el texto como un vector fijo, sin aprovechar la estructura inherente al lenguaje, lo que limitaba su capacidad para tareas complejas de comprensión semántica (Medhat, Hassan y Korashy 2014).

Con la llegada de las redes convolucionales (*CNN*), se generó una mejora significativa, al permitir la extracción automática de patrones locales y n-gramas mediante filtros convolucionales. Como lo enuncia Kim (2014), esta arquitectura es especialmente útil para identificar combinaciones de palabras indicativas de polaridad en oraciones y frases cortas, pudiendo superar a modelos clásicos en la clasificación de sentimientos, y estableciendo un precedente para el uso de convoluciones en NLP y facilitando la extracción automática de patrones sin necesidad de una ingeniería de características manual.

Por su parte, las redes neuronales recurrentes (*RNN*), y especialmente sus variantes LSTM (*Long Short-Term Memory*) y GRU (*Gated Recurrent Unit*), introdujeron la capacidad de procesar secuencias, permitiendo modelar dependencias contextuales y relaciones a largo plazo en el texto, resolviendo problemas como el desvanecimiento del gradiente. Esto permitió una interpretación más precisa del significado en textos extensos o complejos. Estas arquitecturas se convirtieron en el estándar para tareas de secuencias en NLP, incluido el SA y ABSA (Hossain et al. 2020; Do et al. 2023).

Como menciona Wang et al. (2016), los mecanismos de atención permitieron a los modelos neuronales enfocar el procesamiento en palabras o frases particularmente relevantes dentro de una secuencia. Aplicados en combinación con RNN, LSTM o CNN, los mecanismos de atención mejoraron la identificación de aspectos y la interpretación contextual de las opiniones, sentando las bases para los modelos ABSA personalizados desarrollados antes del auge de los *Transformers* (Do et al. 2023).

Finalmente, la aparición de técnicas de representación vectorial como *Word2Vec* (Mikolov et al. 2013) y *GloVe* (Pennington, Socher y Manning 2014) permitió codificar información semántica y sintáctica en vectores densos, reemplazando los enfoques tradicionales basados en bolsas de palabras. Previo a la llegada de modelos pre entrenados como BERT, surgie-

ron modelos ABSA custom”que integraban varias de estas técnicas (LSTM, atención, CNN) para abordar específicamente la segmentación de aspectos y la clasificación de sentimientos a nivel de aspecto (Wang et al. 2016). Estos modelos personalizados lograron resultados notables en benchmarks específicos, aunque seguían dependiendo de un diseño manual y presentaban limitaciones en la captura de dependencias globales (Do et al. 2023).

**Tabla 4:** Métodos de Redes neuronales tradicionales

Referencias	Método	Descripción breve
Medhat, Hassan y Korashy 2014 Do et al. 2023	ANN (Redes Neuronales Artificiales)	Capas densas y profundas, limitadas para texto secuencial. Primeros experimentos de deep learning en PLN.
Kim 2014 Do et al. 2023	CNN (Redes Neuronales Convolucionales)	Detectan patrones locales y n-gramas, útiles para extracción de características en frases cortas.
Do et al. 2023 Hossain et al. 2020	RNN (Redes Neuronales Recurrentes)	Modelan secuencias; útiles para capturar dependencias contextuales.
Do et al. 2023 Hossain et al. 2020	LSTM/GRU	Variantes de RNN que resuelven el problema de desvanecimiento del gradiente; mejor para textos largos y dependencias de largo plazo.
Do et al. 2023 Wang et al. 2016	Atención (Attention)	Permite al modelo enfocarse en partes relevantes de la secuencia; se usa junto a RNN/LSTM o como mecanismo separado.

Referencias	Método	Descripción breve
Mikolov et al. 2013 Pennington, Socher y Manning 2014	Embeddings clásicos	Word2Vec, GloVe, FastText: generan vectores densos y semánticos para representar palabras en espacio vectorial.
Wang et al. 2016 Do et al. 2023	Modelos ABSA "custom"	Arquitecturas diseñadas para ABSA antes de los Transformers, combinan LSTM/CNN con atención y clasificación multicriterio.

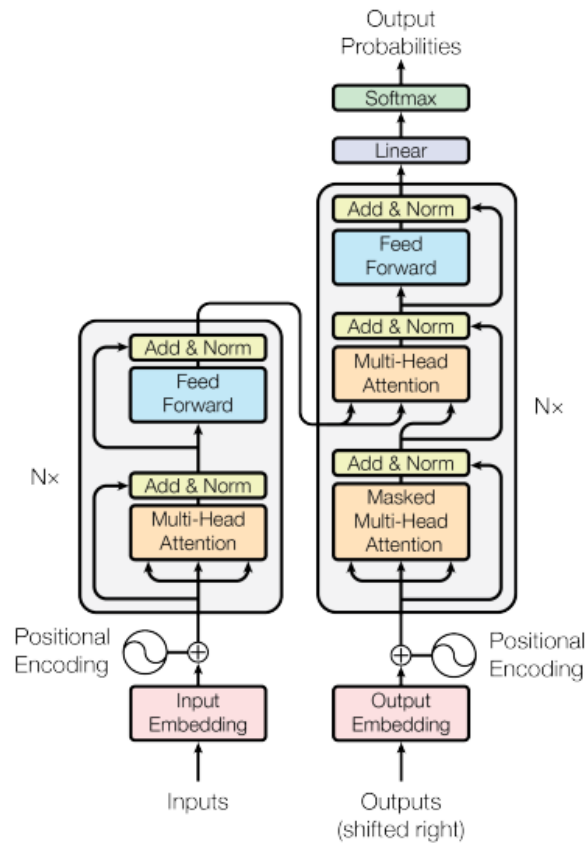
A pesar de estos avances, las arquitecturas basadas en redes neuronales tradicionales presentaban desafíos para modelar dependencias contextuales de largo alcance y para escalar eficientemente en grandes volúmenes de datos. Estas limitaciones motivaron la evolución hacia los modelos *Transformer* y sus derivados, que actualmente dominan el estado del arte en el campo.

### 3.3.2. Transformers, BERT y DeBERTa

En 2017 ocurre un cambio de paradigma en el análisis de texto con la publicación de "Attention Is All You Need" por Vaswani et al. (2017), que introduce la arquitectura Transformer (véase Figura 4). Mientras que las redes neuronales recurrentes y convolucionales habían mejorado sustancialmente la capacidad para modelar texto, seguían presentando limitaciones en el manejo eficiente de dependencias de largo alcance, la paralelización y la escalabilidad. En este contexto, la arquitectura Transformer ha redefinido el estado del arte, permitiendo avances sin precedentes en tareas de comprensión y clasificación de texto (Young et al. 2018).

El modelo Transformer se fundamenta en el mecanismo de autoatención (*self-attention*), que permite que cada elemento de la secuencia (palabra o token) interactúe con todos los demás, ponderando su relevancia contextual. Esta característica elimina la dependencia secuencial de las RNN y posibilita el procesamiento en paralelo, aumentando notablemente la eficiencia y la capacidad de aprendizaje de dependencias globales. El Transformer se compone de bloques de atención multi-cabeza (*multi-head*

*attention*), redes *feed-forward* y codificación posicional, lo que facilita el manejo de secuencias de texto de longitud variable (Vaswani et al. 2017).



**Figura 4:** Arquitectura Transformer. Fuente: Vaswani et al. 2017

En 2019 se produce uno de los avances más notables derivados del modelo Transformer: la introducción de BERT (Bidirectional Encoder Representations from Transformers), presentado por Devlin et al. (2019). BERT está basado únicamente en la parte *encoder* de la arquitectura Transformer y se caracteriza por su modelado bidireccional, que le permite comprender simultáneamente el contexto a la izquierda y derecha de cada palabra. Este preentrenamiento sobre grandes corpus, utilizando tareas como *Masked Language Model* y *Next Sentence Prediction*, ha demostrado mejorar notablemente el rendimiento en múltiples tareas de NLP, incluido el análisis de sentimientos y, especialmente, el ABSA (Young et al. 2018). Además, BERT y sus variantes multilingües han permitido aplicar estos avances a textos en diferentes idiomas, facilitando la transferencia de conocimiento (Do et al. 2023).

Como respuesta a algunas limitaciones de BERT, surgieron modelos mejorados como DeBERTa (Decoding-enhanced BERT with Disentangled Attention), introducido por He et al. (2021). DeBERTa introduce mecanismos como la separación entre las representaciones de contenido y posición (*disentangled*

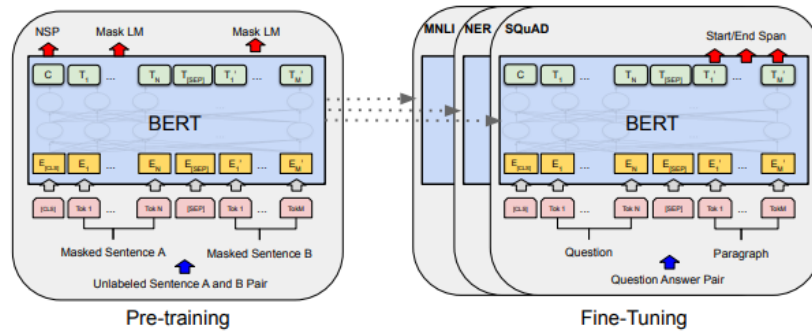


Figura 5: Arquitectura BERT. Fuente: Qian et al. 2022

*attention*), así como mejoras en la decodificación y en los *embeddings* contextuales, como se muestra en la Figura 6. Estas innovaciones permiten una captura más precisa de la semántica y de las relaciones entre palabras, logrando resultados superiores en *benchmarks* de ABSA y otras tareas de comprensión de texto. En el contexto de este trabajo, la selección de un modelo DeBERTa especializado para chino (yangheng/deberta-v3-base-absa-v1.1) responde tanto a la necesidad de gestionar grandes volúmenes de reseñas como a la importancia de captar matices lingüísticos propios de este idioma (Do et al. 2023).

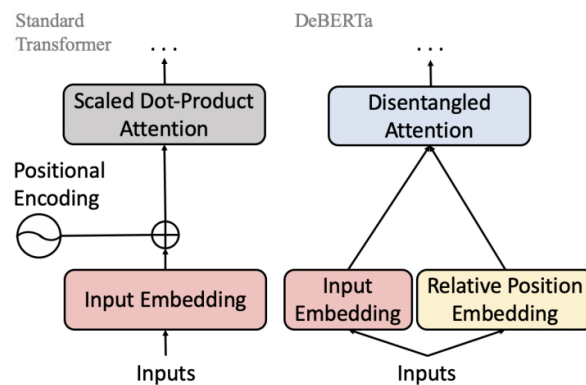


Figura 6: Arquitectura DeBERTa. Fuente: He et al. 2021

Estos modelos preentrenados, como BERT y DeBERTa, se pueden adaptar a tareas específicas mediante *fine-tuning*, como se observa en la Figura 5. Su uso ha impulsado mejoras notables en los *benchmarks* de ABSA y análisis de sentimientos (Do et al. 2023). No obstante, estos modelos presentan desafíos, como la necesidad de grandes recursos computacionales y dificultades de interpretación, lo que sigue motivando avances en el campo (Young et al. 2018).

Los modelos derivados de Transformers han transformado el panorama del NLP, convirtiéndose en el estándar para el SA y ABSA tanto en la academia como en la industria. Su capacidad para procesar

---

texto en paralelo, capturar relaciones complejas y transferir conocimiento a nuevas tareas y dominios justifica su protagonismo en la metodología de este trabajo.

### 3.4. Particularidades del análisis de texto en chino

El análisis de texto en chino presenta una serie de desafíos particulares, siendo el principal la ausencia de espacios para separar palabras, lo que exige técnicas de segmentación léxica especializadas (Jurafsky y Martin 2021). Para abordar la segmentación léxica en chino, se emplea Jieba, una de las librerías más populares y precisas para la tokenización de textos en este idioma (Sun 2012). Esta herramienta facilita dividir oraciones en palabras, paso esencial antes de cualquier análisis basado en características lingüísticas o *embeddings*. Además, la polisemia y ambigüedad léxica de los caracteres incrementa la complejidad del procesamiento, ya que el significado de cada término depende en gran medida del contexto (Liu 2012).

Por otra parte, la limitada disponibilidad de recursos lingüísticos y modelos robustos específicamente diseñados para el chino, como listas de *stopwords* adecuadas o modelos preentrenados sobre grandes corpus, limita las posibilidades de aplicar directamente metodologías desarrolladas para otros idiomas. Esto representa un reto tanto para el análisis de sentimientos convencional como para el ABSA, que exige una comprensión profunda y precisa del texto (Bu et al. 2021). Por estas razones, la adaptación de herramientas y la selección cuidadosa de tecnologías resulta fundamental en trabajos que abordan el análisis automático de reseñas en chino.

---

## 4. Revisión de literatura

En la Sección 4.1 se presenta una revisión de literatura sobre aplicaciones de ABSA en la industria de los restaurantes, priorizando documentos frecuentemente citados y publicados a partir de 2019. En la Sección 4.2 se realiza una revisión de trabajos existentes relacionados que utilizan bases de datos de Dianping.com. Esto para contextualizar la investigación y situar el presente trabajo en el marco de los desarrollos más relevantes del área.

### 4.1. Revisión de literatura industria

Para esta revisión se realizó una búsqueda sistemática en Google Scholar, Scopus y Web of Science, recopilando estudios de ABSA aplicados a la industria de los restaurantes y sitios de reseñas en línea, publicados entre 2019 y 2024. La selección se basó en la frecuencia de citación, rigurosidad académica y relevancia para los objetivos del presente trabajo.

**Tabla 5:** Comparativa de estudios recientes en ABSA en restaurantes

Referencia	Título	Tema	Tecnologías
Krishna et al. (2019)	Sentiment Analysis of Restaurant Reviews Using Machine Learning Techniques	Análisis de sentimiento en reseñas de restaurantes	SVM, Naive Bayes, Random Forest, procesamiento de texto
Hossain et al. (2020)	Sentilstm: a deep learning approach for sentiment analysis of restaurant reviews	Análisis de sentimiento en reseñas de restaurantes	LSTM, deep learning, procesamiento de texto
Asani, Vahdat-Nejad y Sadri (2021)	Restaurant recommender system based on sentiment analysis	Recomendación de restaurantes basada en reseñas online	Minería de opiniones, análisis de sentimiento, sistemas de recomendación, ML

Referencia	Título	Tema	Tecnologías
Jabreel, Ibrahim y Moreno (2021)	Sentiment analysis-based multi-person multi-criteria decision making for restaurant reviews using deep learning and aspect-based sentiment analysis	Toma de decisiones y recomendación de restaurantes usando reseñas	ABSA, CNN, LSTM, multitask learning, deep learning
Pandey, Dey y Jain (2022)	Aspect Based Sentiment Analysis for Explicit and Implicit Aspects in Restaurant Review using Grammatical Rules, Hybrid Approach, and SentiCircle	ABSA en reseñas de restaurantes (aspectos explícitos e implícitos)	Reglas gramaticales, enfoque híbrido, SentiCircle, técnicas estadísticas
Tayal et al. (2023)	Multi-criteria restaurant ranking based on aspect-based sentiment analysis and pythogenic sets	Ranking de restaurantes a partir de reseñas descompuestas por aspectos	ABSA, conjuntos pythogénicos, análisis multicriterio
Li et al. (2023)	Restaurant survival prediction using customer-generated content: An aspect-based sentiment analysis of online reviews	Predicción de supervivencia de restaurantes a partir de reseñas online	ABSA, PLN, modelos predictivos

En la Tabla 5 se muestran los estudios más recientes sobre análisis de sentimientos en la industria de la restauración. Los enfoques clásicos, representados por Krishna et al. (2019), emplearon algoritmos tradicionales de aprendizaje automático, como SVM, Naive Bayes y Random Forest, estableciendo líneas base de comparación. No obstante, dichos métodos presentaban limitaciones para capturar la complejidad semántica y contextual de los textos.

La adopción de modelos de *deep learning*, como el propuesto por Hossain et al. (2020) basado en

---

LSTM, marcó un avance importante en la capacidad de modelar secuencias y patrones contextuales. En estos casos el análisis continuaba siendo a nivel de frase, pero no de aspecto, situación similar al enfoque tomado por Asani, Vahdat-Nejad y Sadri (2021). Esto sugiere una acogida tardía de la industria con respecto al enfoque ABSA que desde Pontiki et al. (2014) se había vuelto el estándar académico.

En Jabreel, Ibrahim y Moreno (2021) y Pandey, Dey y Jain (2022) se representa la transición a un modelo más granular de análisis de sentimientos por aspectos, abordando tanto aspectos explícitos como implícitos mediante técnicas híbridas y *deep learning*. Estos estudios demostraron que el ABSA permite extraer información mucho más útil y detallada para la toma de decisiones, especialmente en sistemas de recomendación como los propuestos por Tayal et al. (2023), quienes integran los resultados de ABSA en modelos multicriterio para personalizar la experiencia del usuario.

En estudios más recientes, como el de Li et al. (2023), se exploran aplicaciones avanzadas, como la predicción de supervivencia de restaurantes a partir del sentimiento detectado en las reseñas. Este tipo de análisis amplía el impacto potencial del análisis de opiniones al combinarse con otras técnicas de modelos predictivos, sugiriendo así la aplicabilidad de estas metodologías para resolver problemáticas reales de la industria de la restauración.

Con miras a continuar con la aplicación de tecnologías modernas y fortalecer el panorama académico, este trabajo se enfoca en la clasificación de sentimiento con modelos BERT, extracción automática de aspectos y análisis ABSA aplicado a la industria de la restauración. Además, integra un análisis estadístico y visual que vincula los resultados de sentimiento y aspectos con las valoraciones tradicionales, ofreciendo así una visión mucho más detallada y práctica de la experiencia del usuario en restaurantes.

## **4.2. Revisión de literatura Dianping.com**

Siguiendo una estructura similar a la revisión anterior, en esta sección se examinan investigaciones representativas que utilizan bases de datos de Dianping.com. Se destacan los retos y oportunidades que plantea este entorno específico para el desarrollo de soluciones avanzadas de NLP, así como las principales tendencias metodológicas y aplicaciones desarrolladas en el contexto de las reseñas de

restaurantes en China.

**Tabla 6:** Comparativa de estudios recientes de Dianping.com

Referencia	Título	Tema	Tecnologías
Zhang et al. (2014a)	Do Users Rate or Review? Boost Phrase-level Sentiment Labeling with Review-level Sentiment Classification	Análisis de sentimiento a nivel de frase en reseñas de Dianping	Clasificación de sentimiento a nivel de reseña y frase, aprendizaje supervisado
Zhang et al. (2014b)	Explicit Factor Models for Explainable Recommendation based on Phrase-level Sentiment Analysis	Recomendaciones explicables en plataformas de reseñas (Dianping) mediante análisis de sentimiento a nivel de frase	Modelos de factores explícitos, análisis de sentimiento a nivel de frase, sistemas de recomendación
Bu et al. (2021)	ASAP: A Chinese review dataset towards aspect category sentiment analysis and rating prediction	Análisis de sentimiento basado en aspectos y predicción de calificación usando reseñas chinas (incluyendo Dianping)	Creación de dataset, ABSA, categorización de aspectos, modelos supervisados y evaluación con deep learning
Brauwers y Frasin-car (2022)	A survey on aspect-based sentiment classification	Revisión exhaustiva sobre ABSA en diferentes dominios, incluyendo datos de Dianping	Revisión de modelos clásicos y deep learning, análisis comparativo de datasets, discusión de retos y tendencias

Referencia	Título	Tema	Tecnologías
Shu et al. (2024)	A Multi-Criteria Decision Support Model for Restaurant Selection Based on Users' Demand Level: The Case of Dianping.com	Selección de restaurantes según el nivel de demanda de los usuarios utilizando datos de Dianping	Modelos de decisión multicriterio, análisis de reseñas de Dianping, integración de preferencias de usuario

Para esta sección se buscó analizar algunos de los trabajos más citados y relevantes que aplican SA a reseñas de restaurantes usando bases de datos de Dianping.com. Los primeros trabajos pioneros, como los de Zhang et al. (2014a) y Zhang et al. (2014b), abrieron camino al aplicar técnicas de análisis de sentimiento a nivel de frase y desarrollar modelos de recomendación explicable a partir de reseñas reales de Dianping. Para la implementación de este trabajo se empleó una versión recortada de la base de datos de Dianping utilizada en dichas investigaciones.

Con el auge de modelos granulares por aspectos, se desarrollaron investigaciones como la de Bu et al. (2021), que facilitó la evaluación objetiva de nuevos métodos ABSA y la comparación entre técnicas clásicas y de deep learning en chino. A su vez, Brauwers y Frasinca (2022) subrayan la importancia de Dianping como benchmark en la literatura internacional de ABSA, dado que sus datos permiten explorar retos únicos asociados al procesamiento del chino y la diversidad de aspectos evaluados por los usuarios.

Más recientemente, el trabajo de Shu et al. (2024) destaca el potencial de Dianping para aplicaciones avanzadas como los sistemas de soporte a la decisión multicriterio, integrando información sobre niveles de demanda de los usuarios y la riqueza de las opiniones disponibles en la plataforma. Estos desarrollos reflejan una tendencia clara hacia la integración de ABSA y modelos de recomendación inteligentes sobre datos reales y complejos.

El presente trabajo aborda varios de los desafíos presentes en la literatura, como la segmentación y desambiguación semántica del chino, la detección automática de aspectos relevantes y la integración efectiva de modelos Transformers en el contexto de Dianping. Por otro lado, los trabajos que aplican ABSA sobre Dianping suelen limitarse a la evaluación de modelos o a la propuesta de sistemas de recomendación, mientras que en este caso se busca profundizar en la relación cuantitativa

---

y visual entre los sentimientos detectados, los aspectos identificados y las métricas tradicionales de valoración, aportando así una perspectiva más integral para la toma de decisiones en la industria de la restauración.

---

## 5. Metodología

En esta sección se detalla el proceso metodológico completo seguido para la realización del presente trabajo. En la Sección 5.1 se describe la preparación y limpieza del conjunto de datos, explicando la selección de variables relevantes, los filtros aplicados y la elección final del restaurante analizado. En la Sección 5.2 se aborda el análisis léxico preliminar, incluyendo la tokenización con la librería Jieba, la eliminación de *stopwords*, el análisis de frecuencias y la generación de visualizaciones exploratorias. En la Sección 5.3 se explica el procedimiento de etiquetado automático del sentimiento general a partir del *rating* numérico, con el objetivo de facilitar el entrenamiento supervisado.

En la Sección 5.4 se presenta el ajuste fino del modelo BERT-Base en chino, detallando la configuración de hiperparámetros y la estrategia de entrenamiento utilizada. En la Sección 5.5 se describe la evaluación del modelo y su aplicación completa sobre el corpus de reseñas, incluyendo la comparación con las etiquetas derivadas de los *ratings*. En la Sección 5.6 se desarrolla el análisis ABSA, explicando la extracción y agrupación de aspectos, así como la implementación del modelo DeBERTa para el análisis por aspecto. Finalmente, en la Sección 5.7 se presenta el análisis cuantitativo y estadístico por aspecto, seguido de un resumen visual que integra todas las etapas del *pipeline* metodológico.

### 5.1. Descripción del conjunto de datos

El análisis se realiza sobre un conjunto de reseñas recopiladas de Dianping.com, una de las plataformas de reseñas de restaurantes más populares en China. El *dataset* inicial incluye un total de 467.455 registros, recopilados entre el 7 de julio de 2004 y el 31 de diciembre de 2016. Cuenta con 294.087 usuarios únicos y 206 restaurantes. El *dataset* se encuentra en chino y será abordado en su idioma original a lo largo de este trabajo.

Inicialmente, se realizó un proceso de limpieza de datos e identificación de variables, lo cual permite obtener un mayor contexto acerca de los datos empleados. Aunque no todas las variables fueron utilizadas en el análisis principal, se presenta a continuación una descripción completa para ofrecer un panorama general del conjunto de datos empleado, como se muestra en la Tabla 7.

**Tabla 7:** Descripción de las variables del conjunto de datos inicial

Variable	Tipo de dato	Descripción
Review_ID	Numérica / Texto	Identificador único de la reseña dentro del dataset.
Merchant	Texto	Nombre del restaurante reseñado.
Rating	Numérica (1-5)	Puntuación global otorgada por el usuario al restaurante.
Score_taste	Numérica (0-4)	Puntuación asignada al sabor/comida.
Score_environment	Numérica (0-4)	Puntuación sobre el ambiente o entorno del restaurante.
Score_service	Numérica (0-4)	Puntuación sobre la calidad del servicio recibido.
Price_per_person	Numérica	Precio promedio por persona en la visita (moneda local).
Time	Fecha/Hora	Momento de publicación de la reseña.
Num_thumbs_up	Numérica	Número de “me gusta” recibidos por la reseña.
Num_response	Numérica	Número de respuestas o comentarios a la reseña.
Content_review	Texto	Texto completo de la reseña.
Reviewer	Texto	Alias o identificador del usuario que escribió la reseña.
Reviewer_value	Numérica / Categórica	Valoración interna o nivel del usuario en la plataforma.
Reviewer_rank	Texto / Categórica	Rango o posición del usuario en la plataforma.
Favorite_foods	Texto	Platos favoritos declarados por el usuario (si está disponible).

---

## 5.2. Limpieza y preprocesamiento de datos

El proceso de limpieza y filtrado inicial de los datos se realizó en SAS, aprovechando su capacidad para manejar grandes volúmenes de información y asegurar la integridad de los registros antes del análisis principal. En esta etapa del proceso, se eliminaron las columnas irrelevantes para el correcto análisis del *dataset*; las variables que se conservaron fueron las siguientes:

- **Merchant:** Nombre del restaurante al que corresponde la reseña.
- **Content\_review:** Texto completo de la reseña.
- **Rating:** Puntuación global otorgada por el usuario al restaurante (escala 1-5).
- **Score\_taste:** Puntuación asignada al sabor/comida (escala 0-4).
- **Score\_environment:** Puntuación sobre el ambiente o entorno (escala 0-4).
- **Score\_service:** Puntuación sobre la calidad del servicio (escala 0-4).

Posteriormente se eliminaron los registros con valores ausentes o inválidos en las columnas seleccionadas. Asimismo, se eliminaron registros con textos vacíos o con el marcador “ ”, que denota ausencia de contenido en la plataforma original.

A continuación, se eliminaron duplicados en función del texto de la reseña. Luego, se llevó a cabo una normalización del texto, eliminando espacios, saltos de línea y tabulaciones, y se filtraron las reseñas que tenían menos de 5 caracteres útiles, garantizando que cada opinión contuviera suficiente información para su análisis.

En la siguiente etapa, se validaron y convirtieron a formato numérico las variables de puntuación, excluyendo cualquier registro con valores fuera de rango.

Finalmente, para el análisis posterior, se seleccionó el restaurante con mayor número de opiniones “Restaurante Chunzai (Sucursal Plaza Guangming)”. Este enfoque permitió garantizar la calidad, homogeneidad y relevancia de la muestra para los modelos y experimentos posteriores, a su vez que permite mejorar la homogeneidad, reducir ruido y la replicabilidad.

**Tabla 8:** Resumen de etapas de limpieza y filtrado del conjunto de datos

<b>Etapas</b>	<b>Número de observaciones</b>
Descarga del dataset original	467,455
Eliminación de registros nulos y textos vacíos	467,447
Eliminación de duplicados	453,825
Limpieza y validación de variables numéricas	453,679
Filtrado por restaurante con más reseñas	19,732

*Nota: La limpieza incluyó la eliminación de columnas irrelevantes, filtrado de registros nulos o vacíos, eliminación de duplicados, validación de rangos para variables numéricas y selección del restaurante con mayor cantidad de opiniones, “Restaurante Chunzai (Sucursal Plaza Guangming)”, para asegurar homogeneidad y calidad en los análisis posteriores.*

Esta Tabla 8 resume el flujo de limpieza, permitiendo visualizar el tamaño de la muestra tras cada filtrado y asegurando la trazabilidad del proceso. Finalmente, tras completar el proceso de limpieza y filtrado en SAS, se exportó un archivo limpio en formato .csv con las columnas relevantes para su posterior tratamiento en Python.

Posteriormente, todos los siguientes pasos se realizaron utilizando Python y las librerías especializadas descritas en secciones posteriores. El código, la base de datos y otra información relevante para la realización de este trabajo se encuentran en el repositorio <https://github.com/julianreina22/TFM>.

### **5.3. Tokenización y análisis léxico preliminar**

Con la base de datos limpia, se procede a realizar la tokenización del texto, empleando la librería Jieba, ampliamente reconocida por su precisión en la segmentación léxica de textos en chino. Posteriormente, para eliminar palabras vacías, se utilizó la lista de *stopwords* en chino disponible en Github Liu (2018), con el objetivo de reducir el ruido y concentrar el análisis en términos informativos y relevantes.

Luego se realizó un análisis de frecuencia de *tokens* que permitió identificar las palabras más comunes, proporcionando una primera visión sobre los temas y aspectos recurrentes en las reseñas. Además,

---

se verificó la correcta eliminación de las *stopwords*.

Finalmente, para facilitar la interpretación visual y dar mayor transparencia al proceso de tokenización, se tradujeron al español los 30 términos más repetidos y se generó una nube de palabras (*wordcloud*) a partir de ellos. Esta visualización, que se presenta en la sección de Resultados (Figura 9), ilustra la frecuencia relativa de los términos y proporciona una primera impresión sobre las temáticas predominantes en las opiniones de los usuarios.

#### **5.4. Etiquetado de sentimientos por *Rating***

Para facilitar la evaluación automatizada y supervisada del sentimiento general de las reseñas, se realizó una etiquetación automática de las observaciones a partir de la variable *Rating*. Considerando la escala de 1 a 5 estrellas, se estableció una regla sencilla: las valoraciones de 4 o 5 se consideran positivas, las de 3 neutras y las de 1 o 2 negativas, como se aprecia en Tabla 14. Esta transformación permite convertir una variable ordinal numérica en una variable categórica nominal de sentimiento (positivo, neutro, negativo), que resulta más adecuada para los modelos de análisis de sentimientos supervisados.

De este modo, la estrategia adoptada resulta especialmente relevante en contextos donde el etiquetado manual completo no es viable, permitiendo que el *rating* actúe como una referencia *proxy* de la percepción del usuario.

#### **5.5. Ajuste fino del modelo BERT**

La elección de este modelo se justifica por su capacidad para capturar relaciones contextuales complejas y matices semánticos del idioma, aspectos cruciales para el éxito del ABSA. La arquitectura bidireccional de BERT permite aprovechar tanto el contexto anterior como el posterior a cada palabra, mejorando la detección de opiniones y la clasificación de polaridad incluso en frases cortas o ambiguas, una situación habitual en reseñas de usuarios.

---

### 5.5.1. Configuración y entrenamiento

Después de transformar los valores numéricos en etiquetas de sentimiento (positivo, neutro y negativo) para facilitar su interpretación por el modelo, se seleccionó como punto de partida el modelo preentrenado BERT-Base, Chinese (*bert-base-chinese*). Este modelo, desarrollado por Google Research y publicado junto con el artículo original de BERT (Devlin et al. 2019), fue entrenado a gran escala sobre datos textuales en chino y se encuentra disponible en la plataforma *Hugging Face Model Hub* (Research 2018).

A nivel metodológico, se llevó a cabo un proceso de ajuste fino (*fine-tuning*) del modelo sobre el conjunto de reseñas, utilizando dichas etiquetas para adaptar la red neuronal a las particularidades del dominio gastronómico y la expresión de sentimiento en chino. El conjunto de datos se dividió en entrenamiento (80 %) y validación (20 %) con el fin de evaluar el desempeño del modelo de forma robusta. Para esta etapa, se implementó una arquitectura de clasificación multiclase con tres etiquetas (positivo, neutro y negativo).

Como se puede ver en la Tabla 9, el entrenamiento se realizó durante 3 épocas, el tamaño de lote se fijó en 8, como equilibrio entre eficiencia computacional y estabilidad en la actualización de parámetros, y la longitud máxima de entrada se estableció en 128 *tokens*, suficiente para cubrir la mayoría de las reseñas sin truncamientos excesivos. Además, se empleó un *learning rate* de  $2 \times 10^{-5}$  y se introdujeron pesos diferenciados en la función de pérdida (2.5, 1.5 y 1.0) para corregir el fuerte desbalance de clases, incentivando al modelo a prestar mayor atención a las categorías menos representadas.

**Tabla 9:** Resumen de hiperparámetros del ajuste fino

Hiperparámetro	Valor	Justificación
Épocas	3	Evita sobreajuste, converge bien en BERT
Tamaño de lote	8	Equilibrio eficiencia/estabilidad
Longitud máxima	128	Cubre la mayoría de reseñas
Learning rate	$2 \times 10^{-5}$	Estándar para BERT
Pesos de pérdida	2.5, 1.5, 1.0	Corrige desbalance de clases

Estos hiperparámetros fueron definidos siguiendo buenas prácticas establecidas en la literatura sobre ajuste fino de modelos Transformer y adaptados específicamente a la naturaleza del corpus gastronómico. Con ello, se busca mejorar la sensibilidad global del clasificador y optimizar la identificación de opiniones menos frecuentes.

---

## 5.6. Evaluación y aplicación del modelo

Una vez completado el ajuste fino del modelo BERT-Base, Chinese, se procedió a su evaluación sobre el conjunto de validación, conformado por el 20% de las reseñas originalmente etiquetadas. Se generaron predicciones para cada observación y se evaluaron mediante métricas estándar para problemas de clasificación multiclase, incluyendo precisión, recall y F1-score, utilizando la función `classification_report` de *Scikit-learn*. Para complementar este análisis, se representó visualmente la matriz de confusión, lo que permitió identificar la proporción de aciertos y errores entre las clases de sentimiento. Estos resultados aportan una visión integral sobre la capacidad del modelo para distinguir entre opiniones negativas, neutras y positivas, y constituyen la base para su aplicación posterior sobre el *corpus* completo.

Una vez validado el desempeño del modelo, se procedió a aplicar el modelo *fine-tuned* sobre la totalidad del *corpus* de reseñas del restaurante analizado. Este proceso consistió en tokenizar y procesar todas las observaciones del conjunto de datos final, utilizando lotes de tamaño adecuado para optimizar la inferencia. El modelo generó para cada reseña una predicción de sentimiento (negativo, neutro o positivo), la cual fue registrada en una nueva variable dentro del `DataFrame`. De este modo, se obtuvo una columna de sentimiento homogénea para todo el *corpus*, que sirve como base para los análisis posteriores, incluyendo la extracción y análisis de aspectos (ABSA).

### 5.6.1. Comparación entre etiquetas por *rating* y predicción BERT

Para evaluar la concordancia entre el sentimiento etiquetado a partir del *rating* numérico (*proxy* de juicio humano) y el sentimiento predicho por el modelo BERT *fine-tuned*, se calculó una matriz de confusión cruzando ambas variables a nivel de reseña. Este análisis permite identificar hasta qué punto el modelo reproduce el criterio humano implícito en las puntuaciones, detectar posibles sesgos sistemáticos y cuantificar el grado de desacuerdo entre ambos enfoques.

Se reportan tanto la matriz de confusión como la tasa de coincidencia global (porcentaje de casos en los que la predicción de BERT coincide con la etiqueta asignada por *rating*). El análisis visual y cuantitativo de esta matriz permite entender mejor los errores típicos del modelo y orientar futuros ajustes o interpretaciones.

---

Los resultados concretos de esta comparación se discuten y visualizan en detalle en la sección de Resultados.

El ajuste fino del modelo BERT para clasificación general de sentimiento se realizó como paso previo al análisis ABSA. Esta decisión se fundamenta en la necesidad de adaptar el modelo a las particularidades del dominio de las reseñas gastronómicas en chino, así como de garantizar una segmentación y etiquetado homogéneo de las opiniones. El proceso permite filtrar opiniones ambiguas, preparar el corpus para una extracción de aspectos más precisa y validar posteriormente la coherencia entre sentimiento global y sentimiento por aspecto, siguiendo las mejores prácticas reportadas en la literatura.

## 5.7. Análisis ABSA

En esta sección se describe la implementación del ABSA, componente central del presente trabajo. A continuación, se presentan las etapas clave del *pipeline* utilizado, incluyendo la extracción y agrupación de aspectos, la aplicación del modelo DeBERTa y el análisis cuantitativo por aspecto.

### 5.7.1. Extracción y agrupación de aspectos

Para identificar los aspectos temáticos predominantes en las reseñas y estructurar el análisis ABSA, se implementó una estrategia de extracción basada en técnicas de procesamiento gramatical para chino. En primer lugar, se utilizó el módulo de análisis de partes de la oración de la librería Jieba (`jieba.posseg`), que permite, para cada palabra, identificar su categoría gramatical mediante etiquetado tipo *Part-of-Speech tagging* (*POS-tagging*).

Se procesó cada reseña del *corpus* tokenizándola y extrayendo aquellos términos clasificados como sustantivos (aquellos cuyo *tag* comienza por “n”), ya que la literatura previa establece que los aspectos relevantes en ABSA suelen corresponder a entidades, atributos o características, y suelen aparecer como sustantivos en el texto (Li et al. 2023). Para evitar la inclusión de términos espurios o poco frecuentes, se impuso además una condición de longitud mínima de dos caracteres por *token*.

Posteriormente, se seleccionaron para análisis únicamente aquellos aspectos con una frecuencia igual

---

o superior a 100 menciones, lo que asegura la robustez estadística y relevancia práctica de los términos retenidos, siguiendo recomendaciones habituales en trabajos similares (Li et al. 2023). El umbral se fijó tras explorar la distribución de frecuencias y evaluar el número de aspectos obtenidos a distintos cortes.

Para una interpretación más estructurada, los términos más frecuentes fueron agrupados manualmente en tres categorías clave de la industria de la restauración (sabor, servicio y ambiente) siguiendo la taxonomía propuesta por la literatura y utilizada habitualmente en la valoración de restaurantes, así como la estructura original de variables del propio *dataset*. Esta categorización facilita la comparación entre el análisis de sentimientos automático y las métricas clásicas, y permite evaluar la consistencia entre métodos.

### **5.7.2. Aplicación del modelo DeBERTa**

Para el análisis de sentimiento basado en aspectos (ABSA), se implementó un *pipeline* automático utilizando modelos de *deep learning* especializados en chino, siguiendo las mejores prácticas y *benchmarks* internacionales del área. En particular, se empleó el modelo preentrenado DeBERTa-v3 para ABSA, disponible en la plataforma *Hugging Face* bajo el identificador `yangheng/deberta-v3-base-absa-v1.1`, el cual ha demostrado un alto desempeño para la tarea de ABSA multiclase en idioma chino (Yang, Zhang y Li 2023) (He et al. 2021).

El modelo DeBERTa se seleccionó por su capacidad mejorada para captar relaciones semánticas complejas y su entrenamiento específico en tareas de ABSA, lo que lo posiciona por encima de alternativas como BERT-base para esta aplicación concreta. Además, su disponibilidad pública en *Hugging Face* asegura la reproducibilidad y trazabilidad del proceso.

### **5.7.3. Análisis cuantitativo y estadístico por aspecto**

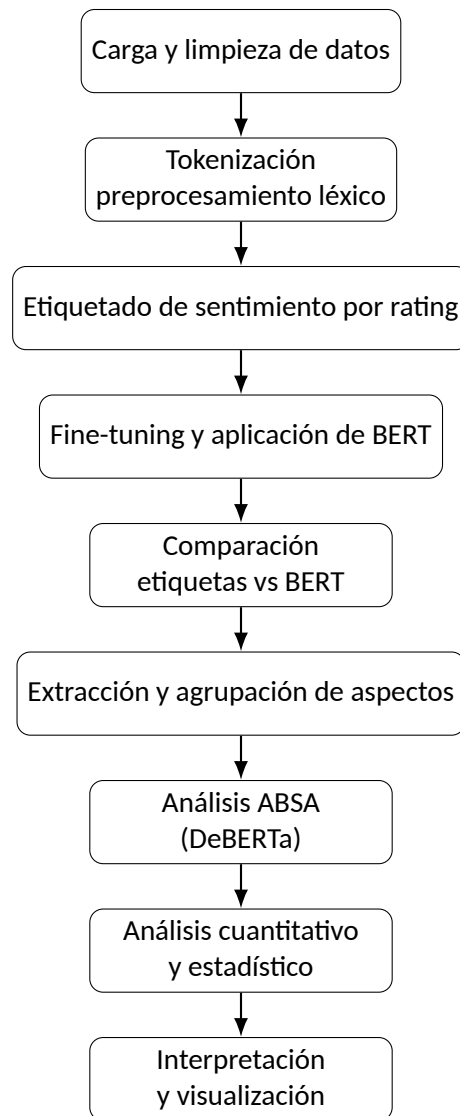
Después de aplicar el modelo ABSA a cada reseña y aspecto, se realiza un análisis cuantitativo y estadístico para evaluar la relación entre el sentimiento detectado automáticamente y las puntuaciones tradicionales asignadas por los usuarios (por ejemplo, `Score_taste` para sabor, `Score_service` para servicio y `Score_environment` para ambiente). Este cruce permite validar la coherencia del

---

modelo con la percepción subjetiva de los usuarios y aporta información adicional para el análisis de la experiencia en restaurantes.

## 5.8. Resumen del *pipeline* metodológico

La Figura 7 sintetiza de forma esquemática el *pipeline* metodológico aplicado en este trabajo, estableciendo la secuencia lógica entre las etapas analíticas y su correspondencia con los fragmentos principales de código descritos en el Anexo.



**Figura 7:** Pipeline metodológico para análisis ABSA sobre reseñas de Dianping. Fuente: Elaboración propia

---

Con esto se completa la descripción detallada de la metodología propuesta para abordar el ABSA en reseñas de restaurantes en chino. Este *pipeline*, diseñado de manera modular y reproducible, sienta las bases para la obtención y validación de resultados robustos. A continuación, en la siguiente sección, se exponen y analizan los principales hallazgos obtenidos a partir de la aplicación de este enfoque metodológico.

---

## 6. Resultados

En esta sección se presentan y analizan en detalle los resultados obtenidos a partir del conjunto final de reseñas seleccionadas. En la Sección 6.1 se realiza un análisis descriptivo del conjunto de datos, destacando las distribuciones generales de puntuaciones y las características principales de las valoraciones. La Sección 6.2 aborda el análisis léxico preliminar, identificando los términos más frecuentes y explorando la semántica general del corpus mediante nubes de palabras. En la Sección 6.3 se presenta la etiquetación inicial de sentimiento a partir del *rating* global, proporcionando una primera aproximación a la polaridad general de las reseñas.

La Sección 6.4 detalla la evaluación del modelo de sentimiento basado en BERT, con sus métricas de desempeño y matriz de confusión. En la Sección 6.5 se realiza una comparación entre las etiquetas de *rating* y las predicciones de BERT, mostrando la identificación y agrupación de términos relevantes. La Sección 6.6 muestra los resultados del modelo DeBERTa para ABSA. Finalmente, la Sección 6.7 profundiza en el análisis cuantitativo y estadístico por aspecto, examinando la relación entre los sentimientos inferidos y las puntuaciones numéricas específicas.

### 6.1. Análisis descriptivo del conjunto de datos

A partir de esta sección, todos los análisis y gráficos se realizaron usando únicamente las reseñas de “Restaurante Chunzai (Sucursal Plaza Guangming)” que, después de la limpieza, son un total de 19.732.

En la Tabla 10 se resumen las estadísticas descriptivas de las principales variables de puntuación del conjunto de datos final, correspondientes a las reseñas del restaurante con mayor número de observaciones. Se observa que la puntuación global (*Rating*) presenta una media de 4.10 y una mediana de 4, con una desviación estándar de 0.95, lo que indica una tendencia general hacia valoraciones positivas y una dispersión moderada dentro del rango permitido (1 a 5). Las valoraciones de sabor (*Score\_taste*), ambiente (*Score\_environment*) y servicio (*Score\_service*) muestran medias ligeramente superiores a 3, con medianas de 3 y desviaciones estándar cercanas a 1, dentro de un rango de (0 a 4). Estos resultados sugieren que, en promedio, las experiencias reportadas por los usuarios son mayoritariamente favorables, aunque existe cierta variabilidad en la percepción de los

---

distintos aspectos evaluados.

**Tabla 10:** Estadísticas descriptivas de las puntuaciones principales del conjunto de datos final

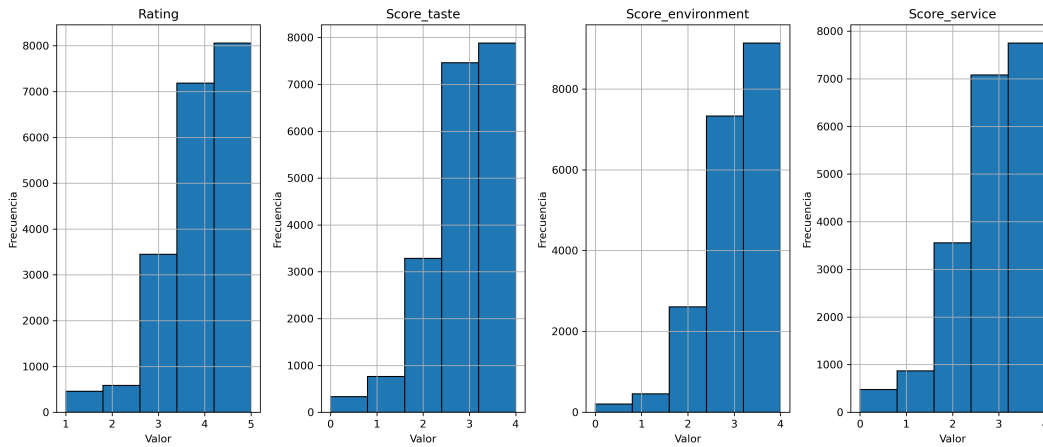
Variable	Media	Mediana	Desviación estándar	Rango
Rating	4.10	4	0.95	1-5
Score_taste	3.10	3	0.93	0-4
Score_environment	3.25	3	0.85	0-4
Score_service	3.05	3	0.98	0-4

Como se puede ver en la Figura 8 y la Tabla 11, la distribución de la puntuación global (Rating) evidencia un claro sesgo hacia las valoraciones más altas, con un predominio de calificaciones de 4 y 5 estrellas, que representan alrededor del 77.2 % del total de reseñas. En contraste, las puntuaciones de 1 y 2 estrellas son considerablemente menos frecuentes, con un 5.3 %, y un 17.5 % como neutras. Esto sugiere que la mayoría de los usuarios tienden a dejar reseñas positivas.

Al analizar los aspectos evaluados Score\_taste, Score\_environment y Score\_service, se observa una dinámica semejante (véase Tabla 12). En el caso de Score\_taste, las valoraciones de 3 y 4 concentran el 77,8 % de las respuestas, mientras que los valores de 0 y 1 apenas alcanzan el 5,6 %. La distribución para Score\_environment es incluso más polarizada; el 83,6 % de las reseñas otorgan puntuaciones de 3 o 4, siendo la puntuación máxima el valor más frecuente. Las puntuaciones de 0 y 1 representan sólo el 3,3 %.

En el caso de Score\_service, aunque las puntuaciones de 3 y 4 siguen predominando, sumando un 75,7 % de las calificaciones, a su vez se observa un ligero aumento de valoraciones de 0 y 1 hasta el 6,8 % del total, lo que indica que la percepción sobre el servicio es ligeramente más variable y presenta más casos de insatisfacción relativa en comparación con los otros dos aspectos.

En síntesis, la distribución de frecuencias de todas las variables de puntuación pone de manifiesto una tendencia marcadamente positiva en las valoraciones emitidas por los usuarios del restaurante analizado, con una concentración clara en los valores máximos y un bajo peso relativo de las experiencias negativas. Este sesgo optimista será considerado en la interpretación y calibración de los análisis posteriores.



**Figura 8:** Histogramas Rating y Scores. Fuente: Elaboración propia.

**Tabla 11:** Frecuencia de la puntuación global (*Rating*) en las reseñas analizadas

Rating	Frecuencia
1	458
2	585
3	3447
4	7186
5	8056

**Tabla 12:** Frecuencia de puntuaciones por aspecto específico en las reseñas analizadas

Valor	Score_taste	Score_environment	Score_service
0	334	201	479
1	764	456	868
2	3287	2610	3558
3	7466	7332	7079
4	7881	9133	7748

*Nota metodológica: Todos los análisis descriptivos y cálculos estadísticos presentados en esta sección se realizaron utilizando Python 3.10, empleando las librerías pandas y numpy para el procesamiento de datos, así como matplotlib para la generación de gráficos.*

---

En conjunto, estos análisis descriptivos permiten caracterizar el comportamiento general de las valoraciones en las reseñas y proporcionan el contexto necesario para los análisis de sentimiento y extracción de aspectos que se abordan en las secciones siguientes.

## 6.2. Análisis léxico preliminar

Se presenta un análisis preliminar del léxico empleado en las reseñas del restaurante seleccionado, considerando únicamente los tokens más frecuentes tras la eliminación de palabras vacías (*stop-words*) del corpus final. Como muestra la Tabla 13, los términos más frecuentes se relacionan de manera directa con la valoración de la comida, tales como “*sabor*”, “*delicioso*” o “*tofu*”, así como con la calidad del servicio, destacando “*servicio*” y “*camarero/a*”. Igualmente, figuran con alta frecuencia palabras asociadas al ambiente y a la experiencia global en el establecimiento.

Es relevante mencionar la aparición de términos vinculados a promociones y dinámicas de consumo, como “*compra grupal*”, así como palabras asociadas a la cantidad, presentación o características específicas de los platos. Se observa además una presencia destacada de expresiones de connotación positiva, como “*no está mal*”, “*recomendar*” o “*especialmente*”, lo que apunta a una tendencia general hacia la evaluación favorable. Sin embargo, la identificación de términos como “*no*”, “*sin*” o “*un poco*” indica la existencia de opiniones matizadas o críticas <sup>1</sup>, que enriquecen la polaridad del corpus.

**Tabla 13:** Tokens más frecuentes (sin stopwords).

Traducción al español	Frecuencia
No está mal	8793
Sabor	7532
Ambiente	5396
Delicioso / Rico	3959
Presentación de los platos / Calidad de los platos servidos	2900
Bastante / Relativamente	2758
Tofu	2467
Servicio	2383

---

<sup>1</sup>Las palabras o expresiones entre comillas y en cursiva han sido traducidas al español con fines de claridad expositiva. En el corpus original, todas las palabras aparecen en chino.

Napoleón (milhojas)	2378
Un poco / Algo	2128
Sensación / Impresión	1917
No / Sin	1688
Camarero/a / Personal de servicio	1643
Gusta / Me gusta	1603
Compra grupal / Oferta grupal	1503
Muchos / Bastante	1466
La próxima vez	1434
Precio	1431
Platos / Tipo de cocina	1413
Esperar mesa / Hacer fila	1388
Sabor / Gusto	1376
Recomendar / Recomendado	1371
Restaurante	1349
Especial / Especialmente	1321
Creer / Opinar / Pensar	1298
Pichón asado / Paloma joven	1222
Uno / Un	1190
Amigo/a	1094
Porción / Cantidad servida	1068
Ganso vegetariano (plato de tofu)	1049

La Figura 9 muestra una nube de palabras generada a partir de los 30 *tokens* más frecuentes<sup>2</sup>, permitiendo visualizar de manera intuitiva la relevancia relativa de cada término dentro del corpus analizado. El tamaño de cada palabra en la nube es proporcional a su frecuencia, facilitando la identificación de los conceptos centrales en las reseñas.

Este análisis léxico preliminar proporciona una visión general de los aspectos que más destacan los usuarios en sus reseñas, lo que resulta útil para orientar análisis más avanzados, como la identificación

<sup>2</sup>Las palabras o expresiones en la nube han sido traducidas al español con fines de claridad expositiva. En el corpus original, todas las palabras aparecen en chino.



---

mente. Este resultado confirma la tendencia previamente identificada hacia valoraciones favorables, evidenciando un sesgo positivo en las experiencias reportadas por los usuarios.

Estos resultados globales constituyen la referencia inicial para el análisis de sentimiento en el conjunto de datos. En las secciones siguientes, se profundizará en el análisis de sentimiento mediante modelos de aprendizaje profundo y se evaluará la correspondencia entre estas etiquetas y las predicciones automáticas generadas por los modelos de lenguaje natural.

## 6.4. Evaluación del modelo de sentimiento

Para evaluar el rendimiento del modelo de clasificación de sentimiento (basado en BERT<sup>4</sup>), se empleó un conjunto de validación compuesto por 3,947 reseñas, obtenido mediante muestreo aleatorio con semilla 42 para garantizar la reproducibilidad. En la Tabla 15 se presenta el reporte de métricas de precisión, exhaustividad (*recall*) y medida F1 para cada clase, así como los valores promedio.

**Tabla 15:** Reporte de clasificación para el modelo de sentimiento (muestra de validación).

Clase	Precisión	Recall	F1-score	Soporte
Negativo	0.38	0.19	0.25	194
Neutro	0.39	0.15	0.21	673
Positivo	0.82	0.95	0.88	3080
<b>Exactitud global</b>	0.78			
<b>Media macro</b>	0.53	0.43	0.45	3947
<b>Media ponderada</b>	0.72	0.78	0.73	3947

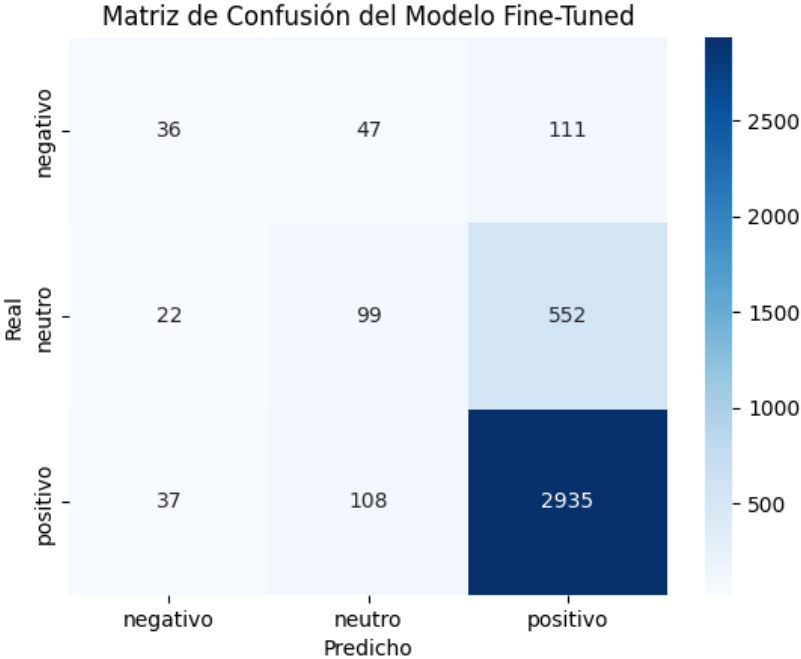
Los resultados muestran que el modelo presenta un rendimiento sobresaliente en la identificación de opiniones positivas (F1=0.88, recall=0.95), pero tiene un desempeño considerablemente menor en las categorías neutro y negativo (medidas F1 de 0.21 y 0.25, respectivamente). La exactitud global alcanza el 78 %, aunque la media macro de F1 es de solo 0.45, lo que evidencia un desequilibrio en la clasificación debido a la fuerte desproporción de clases en el conjunto de datos.

Estos resultados sugieren que, si bien el modelo es adecuado para identificar sentimientos positivos, su capacidad para distinguir opiniones negativas o neutras es limitada. Esto podría deberse tanto al sesgo de la muestra (predominio de reseñas positivas) como a la posible falta de ejemplos suficientes

---

<sup>4</sup>bert-base-chinese

de clases minoritarias para un aprendizaje efectivo del modelo. En consecuencia, los análisis posteriores deberán considerar esta limitación, especialmente al interpretar resultados para clases menos representadas.



**Figura 10:** Matriz de confusión del modelo BERT (muestra de validación). Fuente: Elaboración propia.

La Figura 10 representa la matriz de confusión del modelo de clasificación de sentimiento evaluado sobre la muestra de validación. En esta visualización se pueden observar con claridad los aciertos y errores cometidos por el modelo en cada clase (negativo, neutro y positivo), permitiendo un análisis más detallado de su rendimiento.

En primer lugar, se destaca que el modelo tiene un desempeño muy sólido en la predicción de reseñas positivas. De las 3,080 muestras que realmente pertenecen a esta clase, 2,935 fueron clasificadas correctamente, lo que representa un porcentaje de acierto del 95,3%. Esta alta tasa de aciertos está en línea con el valor de *recall* reportado en el análisis cuantitativo (0.95), y muestra que el modelo ha aprendido de manera efectiva a reconocer esta categoría de sentimiento, probablemente influido por la alta representación de esta clase en los datos de entrenamiento.

En contraste, el rendimiento sobre las clases negativas y neutras es considerablemente más bajo. Por ejemplo, de las 194 reseñas realmente negativas, solo 36 fueron correctamente clasificadas como tales (18,6%). El modelo confundió la mayoría con reseñas positivas (111 casos), y el resto con neu-

---

tras (47 casos). En el caso de las reseñas neutras (673 en total), únicamente 99 fueron identificadas correctamente, mientras que 552 fueron erróneamente clasificadas como positivas y 22 como negativas. Estas confusiones reflejan que el modelo tiende sistemáticamente a sobrepredecir la clase positiva, probablemente como resultado del sesgo de clase en los datos.

Este patrón evidencia un problema de desbalance que afecta la equidad del clasificador. Aunque el modelo alcanza una precisión global del 78 %, esta métrica se ve impulsada principalmente por el buen rendimiento en la clase dominante (positiva). Sin embargo, al observar la matriz de confusión, queda claro que el modelo falla con frecuencia al distinguir entre opiniones neutras y negativas, y presenta una marcada tendencia a agruparlas dentro de la categoría positiva. Esto reduce significativamente la utilidad del modelo en contextos donde sea importante detectar insatisfacción o evaluar matices en la experiencia del cliente.

En resumen, la matriz de confusión complementa el análisis estadístico previo al revelar que el modelo, aunque eficaz para detectar sentimientos positivos, es deficiente en la identificación de emociones más complejas o críticas. Este hallazgo sugiere la necesidad de ajustar el enfoque de entrenamiento en futuras versiones del modelo, mediante técnicas de balanceo de clases o aumento de datos que permitan mejorar la sensibilidad hacia sentimientos menos representados.

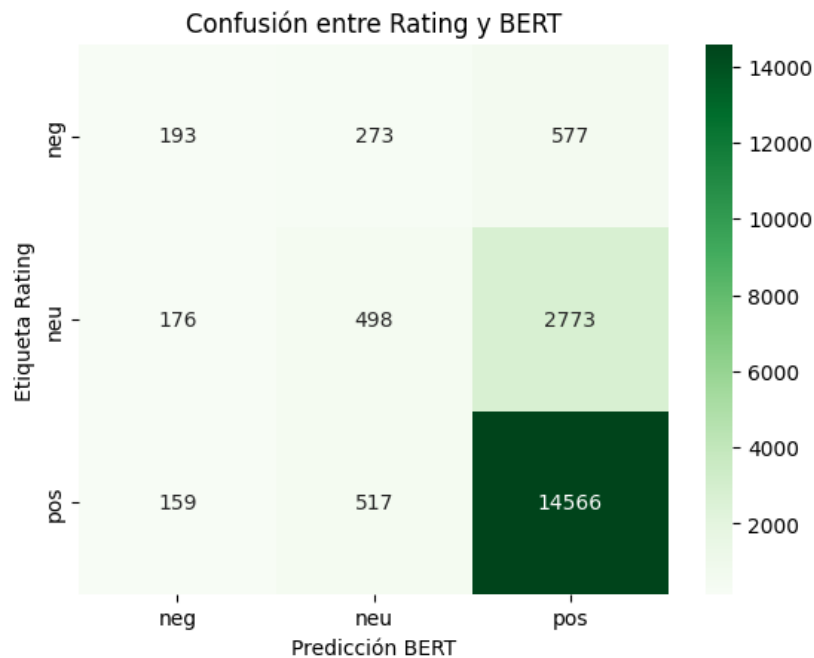
**Tabla 16:** Distribución de sentimientos detectados por BERT en el conjunto completo de reseñas

Sentimiento	Frecuencia	Porcentaje (%)
Positivo	17,916	90.8
Neutro	1,288	6.5
Negativo	528	2.7
<b>Total</b>	<b>19,732</b>	<b>100.0</b>

Posteriormente, se aplicó el modelo BERT al conjunto completo de reseñas (19,732 observaciones) para evaluar la distribución global de sentimiento detectado directamente desde el texto. La Tabla 16 resume estos resultados, mostrando que el 90.8 % de las reseñas fueron clasificadas como positivas, el 6.5 % como neutras y únicamente el 2.7 % como negativas. Esta distribución revela un sesgo aún más pronunciado hacia el sentimiento positivo en comparación con la muestra inicial y refleja la dificultad del modelo para identificar críticas o matices neutros en un contexto altamente polarizado hacia la satisfacción.

## 6.5. Comparación entre etiquetas de *Rating* y predicción BERT

La Figura 11 muestra la matriz de confusión que compara la etiqueta de sentimiento inferida a partir de las puntuaciones explícitas (*Rating*) con la predicción automática realizada por el modelo BERT sobre el texto de la reseña. En total, se evaluaron 19,732 observaciones.



**Figura 11:** Matriz de confusión entre etiquetas de *Rating* y predicciones BERT. Fuente: Elaboración propia.

En primer lugar, se observa que la mayor coincidencia se da en la clase positiva. De las reseñas calificadas como positivas por el *Rating*, 14,566 fueron también clasificadas como positivas por BERT, lo que refleja una concordancia alta en esta clase. Sin embargo, se presentan 517 casos etiquetados como positivos que fueron predichos como neutros, y 159 como negativos, lo que indica cierta variabilidad en la interpretación textual del modelo frente a la puntuación numérica otorgada por los usuarios.

En el caso de las reseñas neutras, la concordancia disminuye significativamente. Solo 498 reseñas neutras según el *Rating* fueron también predichas como neutras por BERT, mientras que una gran parte de las reseñas (2,773) fueron clasificadas como positivas y 176 como negativas. Este patrón sugiere que BERT tiende a agrupar muchas opiniones moderadas hacia la categoría positiva, probablemente influenciado por el sesgo general de positividad presente en el corpus.

---

Respecto a la clase negativa, la concordancia es baja: de las 1,043 reseñas con *Rating* negativo, sólo 193 fueron correctamente identificadas como negativas por BERT. En cambio, 273 fueron predichas como neutras y 577 como positivas. Esto confirma que el modelo tiene dificultades para identificar críticas explícitas, ya sea por la suavidad en la redacción o por el bajo número de ejemplos negativos en los datos de entrenamiento.

En conjunto, la matriz evidencia una tasa de coincidencia general entre *Rating* y BERT del 77,3%. Si bien este nivel de acuerdo puede considerarse aceptable, el análisis detalla que el modelo es particularmente eficaz en la clase positiva, pero menos preciso en las clases neutra y negativa. Este hallazgo refuerza la conclusión de que las reseñas escritas contienen matices y componentes subjetivos que no siempre coinciden con la puntuación explícita asignada por los usuarios.

## 6.6. Análisis de aspectos y resultados ABSA

Tras el análisis global de sentimiento, se realizó un análisis basado en aspectos (*Aspect-Based Sentiment Analysis*, ABSA) para descomponer la polaridad percibida por los usuarios en relación con componentes específicos de la experiencia (por ejemplo, sabor, ambiente y servicio). Para ello, se utilizó el modelo preentrenado `yangheng/deberta-v3-base-absa-v1.1` de Hugging Face.

En primer lugar, se identificaron automáticamente términos frecuentes mediante la segmentación léxica con `jieba`, obteniéndose una lista de los sustantivos más relevantes. Para facilitar el análisis detallado, los aspectos extraídos de las reseñas se agruparon en tres categorías principales: **sabor**, **servicio** y **ambiente**.

1. En la categoría de **sabor** se incluyeron términos relacionados con el gusto y la calidad de los alimentos, como “sabor”, “gusto”, “comida”, “plato”, “textura”, “preparación”, “tofu”, “napoleón”, “tipo de plato”, “pichón”, “ganso vegetariano”, “porción”, “cantidad”, “postre”, “oferta especial”, “mariscos”, “variedad”, “queso”, “té de la tarde”, “durian”, “frescura”, “curry”, “frutas”, “relación calidad-precio” y “cosas”.
2. En **servicio** se agruparon palabras vinculadas a la atención recibida, tales como “servicio”, “camarero”, “actitud del servicio” y “cupón de descuento”.

3. En **ambiente** se consideraron aspectos relativos al entorno y la experiencia en el lugar, incluyendo términos como “ambiente”, “lugar”, “plaza”, “mesa de té” y “luz”.

La Tabla 17 muestra la distribución final de estos aspectos clave detectados. De esta manera, se logró una clasificación más estructurada de los aspectos, lo que permitió analizar con mayor precisión las opiniones de los usuarios en torno a cada dimensión relevante de la experiencia gastronómica.

**Tabla 17:** Distribución de aspectos clave mencionados en las reseñas (traducido al español).

Aspecto	Frecuencia
Sabor	31,758
Ambiente	9,597
Servicio	4,075

Como ejemplo, la Tabla 18 presenta casos ilustrativos obtenidos directamente del modelo ABSA, donde se observa el aspecto identificado, el sentimiento asignado y el nivel de confianza de la predicción, proporcionando una visión concreta de cómo el modelo interpreta los aspectos mencionados en las reseñas.

**Tabla 18:** Ejemplos de resultados ABSA reales (modelo Hugging Face).

Reseña (español)	Aspecto	Sentimiento	Confianza
El ambiente es agradable, el servicio es muy bueno y los platos llegan rápido.	Servicio	Positivo	0.999
El ambiente es bastante bueno y con ambiente romántico, pero la mesa para dos personas es realmente pequeña; si pides mucho, no cabe todo...	Ambiente	Positivo	0.999
El consumo es bastante alto y los alimentos pedidos no se podían pagar con cupón, lo cual fue decepcionante. Los camareros tampoco nos lo informaron.	Servicio	Negativo	0.997

Reseña (español)	Aspecto	Sentimiento	Confianza
La actitud del servicio es mala, la calidad de los platos es mala, hasta la lechuga que pedimos estaba podrida. "Sin fronteras" significa que tienen de todo, pero nada es bueno.	Servicio	Negativo	0.997
Yo también soy empleado de ChunZai (sucursal Shanghái), y lo que más me gusta de nuestra tienda son los postres.	Sabor	Neutro	0.983

*Nota: Las reseñas están traducidas al español y resumidas para facilitar la lectura. Los ejemplos fueron seleccionados manualmente a modo ilustrativo.*

En la Tabla 19 se presenta un ejemplo real con diferente polaridad y presencia de los tres aspectos. La reseña original traducida al español es la siguiente: "El sabor está bien, la actitud del servicio es aceptable, pero no hay nada especialmente destacable en el ambiente." Este caso resulta especialmente ilustrativo, ya que en una única reseña el modelo ha identificado opiniones diferenciadas sobre tres aspectos clave: sabor, servicio y ambiente, asignando a cada uno de ellos una polaridad distinta. Concretamente, el modelo considera la opinión sobre el sabor como neutral ("El sabor está bien"), califica el servicio de forma positiva ("La actitud del servicio es aceptable/buena") y evalúa el ambiente con una polaridad negativa ("El ambiente no tiene nada especialmente destacable").

**Tabla 19:** Ejemplo real de reseña con las tres polaridades en aspectos distintos.

**Reseña global (traducción):** "El sabor está bien, la actitud del servicio es aceptable, pero no hay nada especialmente destacable en el ambiente."

Reseña (español)	Aspecto	Sentimiento	Confianza
El sabor está bien.	Sabor	Neutral	0.489
La actitud del servicio es aceptable/buena.	Servicio	Positivo	0.915
El ambiente no tiene nada especialmente destacable.	Ambiente	Negativo	0.544

*Nota: La reseña está traducida al español y resumida para facilitar la lectura. El ejemplo fue seleccionado manualmente a modo ilustrativo.*

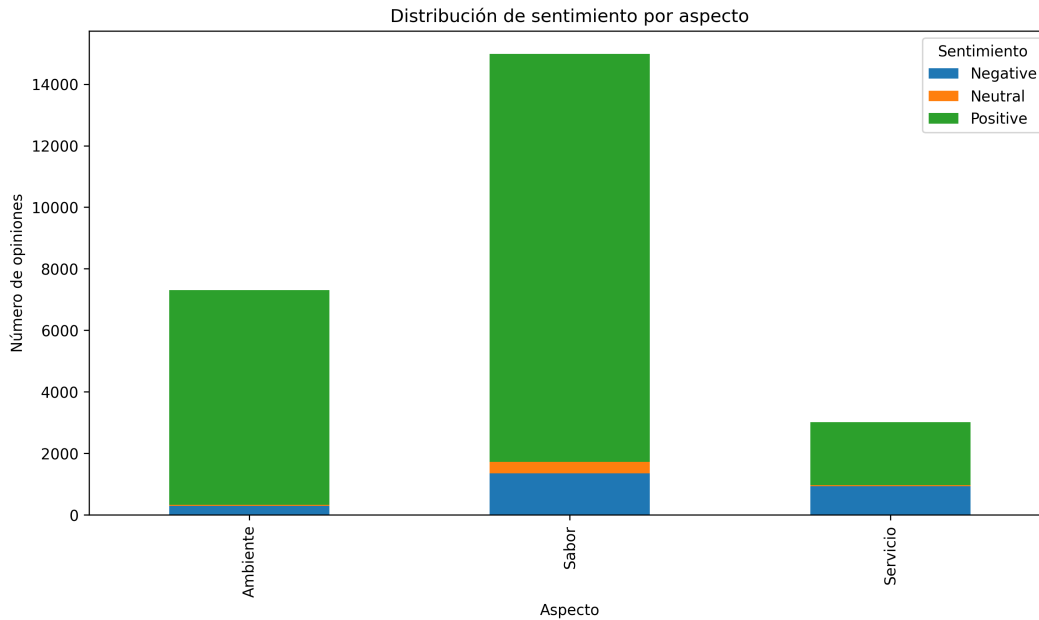
En la Tabla 20 se presenta la distribución completa de sentimientos detectados por el modelo AB-

SA para cada aspecto. Se evidencia una tendencia marcada hacia el sentimiento positivo en las tres categorías.

**Tabla 20:** Frecuencia de sentimientos detectados por aspecto.

Aspecto	Negativo	Neutro	Positivo
Sabor	1,352	377	13,254
Servicio	930	44	2,047
Ambiente	292	40	6,977

La Figura 12 muestra visualmente esta distribución de sentimientos por aspecto en un gráfico de barras apiladas. Se observa que el aspecto 'Sabor' es el más mencionado y también concentra la mayor cantidad de opiniones positivas. El "Servicio" presenta proporcionalmente más críticas negativas, lo cual sugiere que este elemento puede ser un punto de mejora relevante.



**Figura 12:** Distribución de sentimiento por aspecto clave en las reseñas analizadas. Fuente: Elaboración propia.

Finalmente, se exploró la relación entre el sentimiento ABSA y las puntuaciones explícitas otorgadas por los usuarios a cada aspecto (Score\_taste, Score\_service, Score\_environment). La Tabla 21 resume los promedios de estas puntuaciones para cada tipo de sentimiento detectado.

Se aprecia una relación consistente: las reseñas clasificadas como positivas presentan en promedio puntuaciones superiores, mientras que las negativas se asocian a calificaciones bajas. En el caso de

**Tabla 21:** Promedio de puntuación explícita (score) por sentimiento y aspecto.

<b>Aspecto</b>	<b>Negativo</b>	<b>Neutro</b>	<b>Positivo</b>
Sabor	2.32	2.88	3.18
Servicio	1.73	2.77	3.04
Ambiente	2.55	2.93	3.33

”Servicio”, el promedio para sentimiento negativo es especialmente bajo (1.73), lo que refuerza la percepción de este aspecto como el punto más crítico.

En conjunto, el análisis ABSA proporciona una perspectiva detallada y complementaria al análisis global de sentimiento, permitiendo identificar fortalezas y áreas de mejora específicas a nivel de cada aspecto relevante del restaurante. Estos hallazgos motivan el análisis estadístico y cuantitativo detallado que se presenta en la siguiente subsección, donde se profundiza en la relación entre puntuaciones explícitas y sentimiento inferido.

## **6.7. Análisis cuantitativo y estadístico**

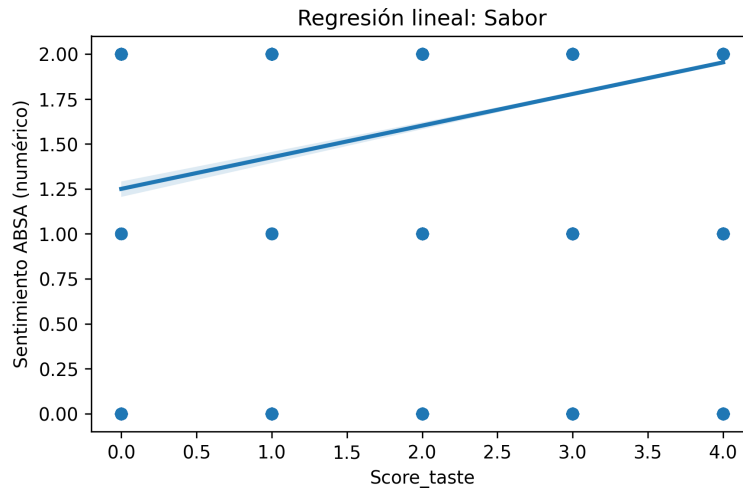
### **Sabor**

En el caso del aspecto *Sabor*, se encontró una correlación de Spearman de 0.22 ( $p < 0.001$ ), lo que indica una relación positiva y estadísticamente significativa entre la puntuación explícita otorgada por los usuarios (*Score\_taste*) y el sentimiento detectado mediante ABSA. El modelo de regresión lineal estimó un coeficiente positivo significativo ( $\text{coef} = 0.18$ ), confirmando que a medida que aumenta la puntuación de sabor, también tiende a incrementarse la probabilidad de que el modelo clasifique el texto con un sentimiento positivo.

El valor de R-cuadrado fue de 0.07, lo que implica que la puntuación explica solo una pequeña parte de la variabilidad del sentimiento ABSA, sugiriendo que otros elementos textuales y contextuales también influyen en la percepción. A pesar de esta baja capacidad explicativa, la pendiente positiva evidencia una coherencia general entre la evaluación explícita y el análisis basado en aspectos.

La Figura 13 ilustra la relación observada entre *Score\_taste* y el sentimiento ABSA en formato numérico (0 = negativo, 1 = neutro, 2 = positivo), donde se aprecia una tendencia ascendente coherente

con los resultados estadísticos.



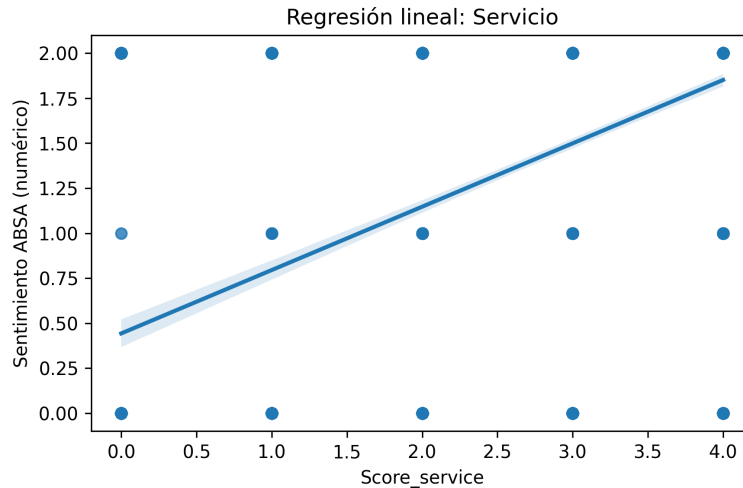
**Figura 13:** Relación entre la puntuación de sabor y el sentimiento ABSA. Fuente: Elaboración propia.

## Servicio

Para el aspecto *Servicio*, se obtuvo una correlación de Spearman de 0.46 ( $p < 0.001$ ), lo que indica una relación positiva y moderadamente fuerte entre la puntuación explícita (*Score\_service*) y el sentimiento inferido por ABSA. El modelo de regresión lineal muestra un coeficiente positivo significativo ( $\text{coef} = 0.35$ ), confirmando que puntuaciones más altas en servicio tienden a asociarse con sentimientos positivos en el texto.

El valor de R-cuadrado alcanzó 0.23, indicando que la puntuación explícita explica aproximadamente el 23% de la variabilidad en el sentimiento ABSA. Este resultado sugiere que el servicio es un factor determinante en la percepción general expresada en las reseñas, más fuerte que en el caso del sabor.

La Figura 14 ilustra esta relación positiva, donde se observa una tendencia ascendente clara y un ajuste lineal más marcado en comparación con otros aspectos.



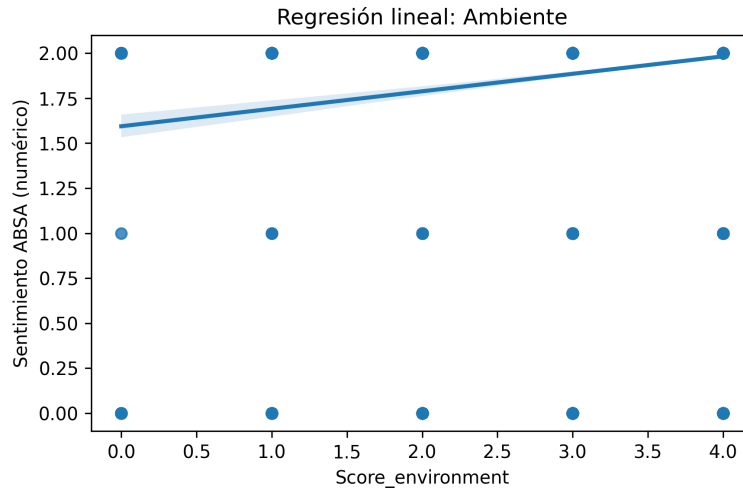
**Figura 14:** Relación entre la puntuación de servicio y el sentimiento ABSA. Fuente: Elaboración propia.

## Ambiente

Para el aspecto *Ambiente*, se obtuvo una correlación de Spearman de 0.16 ( $p < 0.001$ ), lo que indica una relación positiva pero débil entre la puntuación explícita (*Score\_environment*) y el sentimiento inferido por ABSA. El modelo de regresión lineal muestra un coeficiente positivo significativo (coef = 0.10), confirmando que a mayores puntuaciones de ambiente tiende a asociarse un sentimiento más positivo en el texto.

Sin embargo, el valor de R-cuadrado fue bajo (0.04), lo que sugiere que la puntuación explícita de ambiente explica solo el 4% de la variabilidad en el sentimiento ABSA. Este resultado implica que el ambiente podría ser percibido de forma más subjetiva o influenciado por factores adicionales no capturados en la puntuación explícita.

La Figura 15 muestra la relación ascendente entre Score y sentimiento, aunque con una pendiente menos pronunciada y mayor dispersión en comparación con los otros aspectos.



**Figura 15:** Relación entre la puntuación de ambiente y el sentimiento ABSA. Fuente: Elaboración propia.

Estos hallazgos cuantitativos servirán de base para la discusión crítica de los resultados y sus implicaciones en el contexto de la percepción de la experiencia gastronómica, abordada en la sección siguiente.

---

## 7. Discusión

El presente trabajo ha permitido explorar en profundidad la relación entre las puntuaciones explícitas otorgadas por los usuarios (scores numéricos) y los sentimientos inferidos automáticamente mediante modelos de lenguaje, específicamente a través del análisis basado en aspectos (ABSA). Los resultados revelan varios hallazgos relevantes.

En primer lugar, se confirma una tendencia general positiva en las reseñas analizadas, tanto a nivel global, con más del 90 % de predicciones positivas usando BERT, como a nivel de aspectos concretos, donde predominan ampliamente los sentimientos positivos. Este patrón respalda la hipótesis inicial sobre el sesgo positivo presente en plataformas de reseñas, donde los usuarios tienden a compartir principalmente experiencias favorables.

El análisis estadístico detallado mostró que la concordancia entre las puntuaciones numéricas y el sentimiento ABSA es estadísticamente significativa para los tres aspectos analizados (sabor, servicio y ambiente), aunque con magnitudes distintas. El servicio evidenció la mayor asociación (correlación de Spearman = 0.485 y  $R^2 = 0,246$ ), lo que sugiere que los usuarios son más consistentes al reflejar su percepción del servicio tanto en la puntuación como en el texto. En cambio, el sabor y el ambiente presentaron asociaciones más débiles (correlaciones de 0.242 y 0.208, respectivamente), lo que indica mayor subjetividad y matices en estas dimensiones.

Estos hallazgos destacan el valor añadido del ABSA, ya que permite capturar percepciones y matices que no siempre quedan reflejados en las puntuaciones numéricas. De hecho, las puntuaciones tienden a simplificar la experiencia a un solo valor, mientras que el texto aporta riqueza contextual y detalles adicionales que ayudan a comprender mejor la satisfacción del cliente. Este punto coincide con la visión de Liu (2012), quien enfatiza la importancia de integrar fuentes explícitas y textuales para lograr una comprensión más completa de las percepciones reales. Asimismo, se alinea con trabajos recientes que revisan la utilidad del ABSA en contextos como la restauración y la hostelería (Xu et al. 2020; Brauwers y Frasincar 2022).

En conjunto, los resultados destacan el potencial del análisis ABSA como herramienta complementaria para el estudio de la satisfacción del cliente, proporcionando una visión más completa y granular que la basada únicamente en puntuaciones numéricas.

---

## 8. Limitaciones

Si bien los resultados obtenidos son robustos y consistentes con la literatura previa, el presente trabajo presenta varias limitaciones que deben considerarse al interpretar los hallazgos.

En primer lugar, el sesgo de positividad inherente a las reseñas recogidas constituye una restricción importante. La fuerte predominancia de opiniones positivas dificulta la identificación y el modelado correctos de sentimientos negativos y neutros, tanto en los análisis globales como en los específicos por aspecto (ABSA). Este desequilibrio afecta la sensibilidad del modelo, reduciendo su capacidad para detectar matices críticos o experiencias insatisfactorias.

En segundo lugar, la naturaleza subjetiva del lenguaje introduce un grado considerable de variabilidad. Aunque se implementaron procesos de traducción y limpieza exhaustivos, las expresiones idiomáticas, el contexto cultural y el estilo personal de los usuarios pueden generar discrepancias entre la puntuación numérica (score) y el sentimiento textual interpretado por los modelos. Además, el proceso de traducción del chino al español puede haber introducido ambigüedades y pérdida de matices semánticos importantes, afectando la fidelidad del análisis (Jurafsky y Martin 2021; Bird, Klein y Loper 2009).

Otra limitación técnica se relaciona con el uso de modelos preentrenados generales (BERT y DeBERTa) que no han sido específicamente adaptados al dominio gastronómico ni al contexto cultural local, lo que podría introducir sesgos adicionales y reducir la precisión en la identificación de aspectos y sentimientos (Devlin et al. 2019; Vaswani et al. 2017). Asimismo, la validación del modelo se basó únicamente en métricas automáticas (por ejemplo, matriz de confusión y medidas F1), sin incorporar una revisión manual exhaustiva (human-in-the-loop), lo que habría permitido contrastar de forma más rigurosa las etiquetas generadas automáticamente y fortalecer la fiabilidad de los resultados (Liu 2012).

Además, la agrupación de aspectos se realizó en categorías amplias (sabor, servicio y ambiente), sin descomponer en subaspectos más específicos (por ejemplo, presentación de platos, rapidez del servicio o limpieza), lo que puede limitar la capacidad del análisis para captar matices finos (Pontiki et al. 2014; Xu et al. 2020). Por otro lado, no se modelaron explícitamente las posibles dependencias entre aspectos, es decir, cómo la valoración negativa de un aspecto podría influir en la percepción de otro

---

(cross-aspect influence), una línea de análisis relevante sugerida en la literatura reciente (Brauwerters y Frasincar 2022; Zhang et al. 2022).

Finalmente, este estudio se centra en un único restaurante y en un contexto geográfico específico (Guangzhou), lo que limita la generalización de los resultados a otras regiones, tipos de cocina o modelos de negocio. Para incrementar la validez externa, sería recomendable replicar la metodología en muestras más amplias y diversas.

En resumen, estas limitaciones abren líneas de mejora y oportunidades para futuros trabajos, reforzando la importancia de seguir investigando la integración de análisis textual y cuantitativo en el estudio de la satisfacción del cliente.

---

## 9. Conclusiones

El presente trabajo tuvo como objetivo principal aplicar técnicas avanzadas de análisis de sentimiento y extracción de aspectos (ABSA) a reseñas textuales de un restaurante específico en la plataforma Dianping. A partir de un conjunto amplio y real de datos, se desarrolló un pipeline robusto que permitió analizar tanto la polaridad global como el detalle por aspectos clave (sabor, servicio y ambiente).

Los resultados principales muestran una clara tendencia positiva en la mayoría de las reseñas, con más del 90 % de opiniones favorables según el modelo BERT. Sin embargo, la comparación con las puntuaciones explícitas reveló discrepancias significativas, especialmente en las categorías neutra y negativa, donde el modelo presentó dificultades para discriminar correctamente. Este hallazgo refleja tanto el sesgo inherente en los datos como las limitaciones de los modelos actuales en contextos con fuerte desbalance de clases.

El análisis por aspecto mediante ABSA permitió profundizar en la percepción granular de los clientes. Se observaron correlaciones estadísticamente significativas entre las puntuaciones numéricas y los sentimientos inferidos, especialmente en el caso del servicio, que mostró la mayor coherencia (Spearman = 0.49,  $R^2 = 0,25$ ). Por el contrario, sabor y ambiente evidenciaron relaciones más débiles, lo que pone de manifiesto la riqueza subjetiva de las experiencias culinarias y ambientales.

Metodológicamente, este trabajo demuestra el potencial del ABSA como herramienta complementaria al análisis de puntuaciones, capaz de captar matices y percepciones no reflejadas explícitamente en las valoraciones numéricas. La combinación con técnicas estadísticas permitió validar cuantitativamente los resultados, aportando rigor y solidez al análisis.

En conjunto, estos hallazgos destacan la utilidad de enfoques ABSA frente al análisis de sentimiento tradicional, al capturar la complejidad y los matices de la experiencia del cliente en distintas dimensiones. Además, el nivel de confianza asociado a cada predicción ofrece un elemento adicional de robustez para futuros análisis cuantitativos o cualitativos.

---

## 10. Trabajos futuros

A partir de los resultados obtenidos y las limitaciones identificadas, surgen varias líneas de investigación y mejoras metodológicas que pueden explorarse en futuros trabajos.

Se propone ampliar la muestra analizada, incorporando reseñas de múltiples restaurantes, diferentes estilos de cocina y ubicaciones geográficas diversas. Esto permitiría evaluar la robustez y la capacidad de generalización del modelo ABSA, así como contrastar posibles diferencias culturales o contextuales en la percepción del servicio y la comida, tal como se sugiere en estudios recientes en turismo y restauración (Li et al. 2023; Shu et al. 2024).

Resulta relevante abordar el problema del desbalance de clases mediante técnicas específicas. Entre ellas destacan el uso de métodos de *oversampling* (por ejemplo, SMOTE) o *undersampling* controlado, y la aplicación de pérdidas ponderadas durante el entrenamiento de los modelos de lenguaje. Estas estrategias podrían mejorar la detección de opiniones negativas y neutras, contribuyendo a un análisis de sentimiento más equitativo y sensible, en línea con lo expuesto en revisiones generales de análisis de sentimiento (Medhat, Hassan y Korashy 2014; Cambria et al. 2017).

Además, se plantea la exploración de modelos no lineales o híbridos que permitan capturar relaciones más complejas entre los aspectos cuantitativos (scores) y el sentimiento textual. Modelos basados en árboles de decisión, *boosting* (como XGBoost o LightGBM) o redes neuronales profundas podrían ofrecer mejoras significativas en la capacidad predictiva y explicativa, tal como se discute en revisiones recientes sobre NLP y ABSA (Young et al. 2018; Zhang et al. 2022).

Por otra parte, sería interesante incorporar análisis temporales o longitudinales, investigando la evolución del sentimiento y las valoraciones a lo largo del tiempo. Este enfoque permitiría detectar cambios en la percepción del cliente tras eventos específicos (por ejemplo, renovaciones, cambios en el menú o promociones) y anticipar tendencias o posibles crisis reputacionales, como se explora en estudios aplicados a la restauración (Li et al. 2023).

En el futuro, convendría profundizar en la modelización explícita de dependencias entre aspectos (*cross-aspect influence*), para estudiar cómo la percepción negativa en un aspecto podría condicionar la valoración de otros, tal como se plantea en trabajos recientes sobre ABSA (Brauwerts y Frasincar 2022; Zhang et al. 2022).

---

Por último, se sugiere el desarrollo de *dashboards* interactivos para la visualización dinámica de resultados, lo que facilitaría la interpretación por parte de equipos de gestión y marketing. Esta línea contribuiría a un uso más estratégico de los hallazgos, acercando la investigación académica a la toma de decisiones empresariales (Asani, Vahdat-Nejad y Sadri 2021).

En conjunto, estas líneas de trabajo futuro no solo fortalecerían la validez y aplicabilidad del presente estudio, sino que también abrirían nuevas oportunidades para mejorar la comprensión integral de la experiencia del cliente en el sector gastronómico.

---

## 11. Bibliografía

- Asani, E., H. Vahdat-Nejad y J. Sadri (2021). «Restaurant recommender system based on sentiment analysis». En: *Machine Learning with Applications* 6, pág. 100114.
- Bird, Steven, Ewan Klein y Edward Loper (2009). *Natural Language Processing with Python*. O'Reilly Media. ISBN: 978-0596516499.
- Brauwera, G. y F. Frasincar (2022). «A survey on aspect-based sentiment classification». En: *ACM Computing Surveys* 55.4, págs. 1-37.
- Bu, J. et al. (2021). «ASAP: A Chinese review dataset towards aspect category sentiment analysis and rating prediction». En: *arXiv preprint arXiv:2103.06605*.
- Cambria, Erik (2017). «Sentiment analysis is a big suitcase». En: *IEEE Intelligent Systems* 32.6, págs. 136-144.
- Cambria, Erik et al. (2017). «Sentiment Analysis: The State of the Art and Future Directions». En: *Information Fusion* 36, págs. 1-11. DOI: 10.1016/j.inffus.2017.02.006.
- Devlin, Jacob et al. (2019). «BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding». En: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Association for Computational Linguistics, págs. 4171-4186. URL: <https://www.aclweb.org/anthology/N19-1423>.
- dianping.com (2022). *Merchant/review/other related FAQ's*. <https://www.dianping.com/help/faq/4#Q49>.
- Do, Hai Ha et al. (2023). «Deep Learning for Aspect-Based Sentiment Analysis: A Comparative Review». En: *IEEE Access* 11, págs. 28563-28584. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0957417418306456>.
- He, Pengcheng et al. (2021). «DeBERTa: Decoding-enhanced BERT with Disentangled Attention». En: *Proceedings of the International Conference on Learning Representations (ICLR)*. URL: <https://arxiv.org/abs/2006.03654>.
- Hossain, E. et al. (2020). «Sentilstm: a deep learning approach for sentiment analysis of restaurant reviews». En: *International Conference on Hybrid Intelligent Systems*. Springer International Publishing, págs. 193-203.

- 
- Hu, Minqing y Bing Liu (2004). «Mining and summarizing customer reviews». En: *Proceedings of the 10th ACM SIGKDD international conference on Knowledge discovery and data mining*, págs. 168-177.
- Jabreel, M., M. Ibrahim y A. Moreno (2021). «Sentiment analysis-based multi-person multi-criteria decision making for restaurant reviews using deep learning and aspect-based sentiment analysis». En: *Information Processing & Management* 58.2, pág. 102445.
- Jurafsky, Daniel y James H. Martin (2021). *Speech and Language Processing*. 3rd. Draft, available at <https://web.stanford.edu/~jurafsky/slp3/>.
- Kim, Yoon (2014). «Convolutional Neural Networks for Sentence Classification». En: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, págs. 1746-1751.
- Krishna, A. et al. (2019). «Sentiment Analysis of Restaurant Reviews Using Machine Learning Techniques». En: *Emerging Research in Electronics, Computer Science and Technology: Proceedings of International Conference, ICERECT 2018*. Springer Singapore, págs. 687-696.
- Li, H. et al. (2023). «Restaurant survival prediction using customer-generated content: An aspect-based sentiment analysis of online reviews». En: *Tourism Management* 96, pág. 104707.
- Liu, Bing (2012). *Sentiment Analysis and Opinion Mining*. Vol. 5. Synthesis Lectures on Human Language Technologies 1. Morgan & Claypool Publishers. ISBN: 978-1608458844. URL: <https://www.cs.uic.edu/~liub/FBS/SentimentAnalysis-and-OpinionMining.pdf>.
- Liu, Chao-Hong (2018). *HIT stopwords - Chinese stopwords list*. <https://github.com/goto456/stopwords>. Accessed: 2025-06-21.
- Medhat, Walaa, Ahmed Hassan y Hoda Korashy (2014). «Sentiment analysis algorithms and applications: A survey». En: *Ain Shams Engineering Journal* 5.4, págs. 1093-1113. DOI: 10.1016/j.asej.2014.04.011.
- Mikolov, Tomas et al. (2013). «Efficient estimation of word representations in vector space». En: *arXiv preprint arXiv:1301.3781*.
- Pandey, H., S. Dey y S. Jain (2022). «Aspect Based Sentiment Analysis for Explicit and Implicit Aspects in Restaurant Review using Grammatical Rules, Hybrid Approach, and SentiCircle». En: *Procedia Computer Science*. Vol. 218, págs. 1449-1458.

- 
- Pang, Bo, Lillian Lee y Shivakumar Vaithyanathan (2002). «Thumbs up? Sentiment classification using machine learning techniques». En: *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, págs. 79-86.
- Pennington, Jeffrey, Richard Socher y Christopher D Manning (2014). «Glove: Global vectors for word representation». En: *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, págs. 1532-1543.
- Pontiki, Maria et al. (2014). «SemEval-2014 Task 4: Aspect Based Sentiment Analysis». En: *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. Association for Computational Linguistics, págs. 27-35. DOI: 10.3115/v1/S14-2004.
- Pozzi, Federico Alberto et al. (2016). *Sentiment Analysis in Social Networks*. Morgan Kaufmann.
- Qian, Jing et al. (2022). «Limitations of Language Models in Arithmetic and Symbolic Induction». En: *arXiv preprint arXiv:2208.05051*. DOI: 10.48550/arXiv.2208.05051. URL: <https://arxiv.org/abs/2208.05051>.
- Research, Google (2018). *BERT-Base, Chinese*. <https://huggingface.co/bert-base-chinese>.
- Shu, Z. et al. (2024). «A Multi-Criteria Decision Support Model for Restaurant Selection Based on Users' Demand Level: The Case of Dianping.com». En: *Information Processing Management* 61.3, pág. 103650.
- Sun, Junyi (2012). *Jieba Chinese text segmentation: built to be the best Python Chinese word segmentation module*. <https://github.com/fxsjy/jieba>.
- Tayal, A. et al. (2023). «Multi-criteria restaurant ranking based on aspect-based sentiment analysis and pythogenic sets». En: *Expert Systems with Applications* 215, pág. 119218.
- Turney, Peter D. (2002). «Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews». En: *Proceedings of the 40th annual meeting on association for computational linguistics (ACL)*, págs. 417-424.
- Vaswani, Ashish et al. (2017). «Attention is All You Need». En: *Proceedings of the 31st Conference on Neural Information Processing Systems (NeurIPS 2017)*, págs. 5998-6008. URL: <https://arxiv.org/abs/1706.03762>.

- 
- Wang, Yequan et al. (2016). «Attention-based LSTM for aspect-level sentiment classification». En: *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, págs. 606-615.
- Xu, Hua et al. (2020). «Aspect-Based Sentiment Analysis with Deep Learning: A Survey». En: *IEEE Access* 7, págs. 163677-163693. URL: <https://ieeexplore.ieee.org/document/9260162>.
- Yang, Heng, Chen Zhang y Ke Li (2023). «PyABSA: A Modularized Framework for Reproducible Aspect-based Sentiment Analysis». En: *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, CIKM 2023, Birmingham, United Kingdom, October 21-25, 2023*. Ed. por Ingo Frommholz et al. ACM, págs. 5117-5122. DOI: 10.1145/3583780.3614752. URL: <https://doi.org/10.1145/3583780.3614752>.
- Yang, Heng et al. (2021). «Back to Reality: Leveraging Pattern-driven Modeling to Enable Affordable Sentiment Dependency Learning». En: *CoRR* abs/2110.08604. arXiv: 2110.08604. URL: <https://arxiv.org/abs/2110.08604>.
- Young, Tom et al. (2018). «Recent Trends in Deep Learning Based Natural Language Processing». En: *IEEE Computational Intelligence Magazine* 13.3, págs. 55-75. DOI: 10.1109/MCI.2018.2840738.
- Zhang, Wenxuan et al. (2022). «A Survey on Aspect-Based Sentiment Analysis: Tasks, Methods, and Challenges». En: *IEEE Transactions on Knowledge and Data Engineering*. URL: <https://arxiv.org/abs/2203.01054>.
- Zhang, Yongfeng et al. (2014a). «Do Users Rate or Review? Boost Phrase-level Sentiment Labeling with Review-level Sentiment Classification». En: *Proceedings of the 37th Annual International ACM SIGIR Conference on Research and Development on Information Retrieval (SIGIR 2014)*. Gold Coast, Australia, págs. 1027-1030.
- Zhang, Yongfeng et al. (2014b). «Explicit Factor Models for Explainable Recommendation based on Phrase-level Sentiment Analysis». En: *Proceedings of the 37th Annual International ACM SIGIR Conference on Research and Development on Information Retrieval (SIGIR 2014)*. Gold Coast, Australia, págs. 83-92.