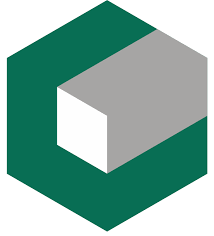
**Universidad Tecnológica de La Habana**

**“José Antonio Echeverría”**

**CUJAE**



**Facultad de Ingeniería Informática**

**Aplicación de modelos *Transformers* al Análisis de Sentimientos en opiniones en español basado en aspectos**

**Informe de prácticas profesional 2**

**Autor**: Miguel Angel Rivero Tapia

**Tutor**: Dr. Alfredo Simón Cuevas

**La Habana, Cuba**

**Septiembre 2024**

**Resumen**

En la actualidad, el auge de la información textual disponible en línea, proveniente de fuentes como noticias, reseñas de productos y redes sociales, ha incrementado la necesidad de contar con métodos automáticos para analizarla de manera rápida y precisa. Este análisis es crucial para la toma de decisiones, por lo que se destinan grandes esfuerzos a identificar el sentimiento de las opiniones en estos textos no estructurados. Sin embargo, en el contexto actual, donde las opiniones se basan en aspectos específicos de un producto o servicio, se ha vuelto esencial ir más allá del análisis general y analizar los sentimientos asociados a cada aspecto individual.

El análisis de sentimientos basado en aspectos se centra en comprender el sentimiento de los usuarios en opiniones textuales sobre características específicas de una entidad. Sin embargo, la identificación de aspectos y su polaridad constituye un desafío, especialmente en el idioma español. La mayoría de las soluciones existentes para esta tarea se han desarrollado y evaluado principalmente en inglés, lo que presenta una menor precisión para el español.

Esta tesis propone un nuevo enfoque para el análisis de sentimientos basado en aspectos, utilizando ensambles de *Transformers*. La idea central consiste en combinar la potencia de los *Transformers* con la técnica de ensambles de modelos, para aprovechar las ventajas de diferentes arquitecturas y estrategias de aprendizaje. Este método consta de tres etapas fundamentales: preprocesamiento del texto, extracción de los aspectos y predicción de polaridad. La solución se evaluó utilizando el conjunto de datos de *SemEval* 2016 *Task* 5, específicamente diseñado para el análisis de sentimientos basado en aspectos.

**Palabras claves:** Aprendizaje profundo, *Transformers*, análisis de sentimientos, extracción de aspectos, ensamble.

**Abstract**

Nowadays, the rise of textual information available online, coming from sources such as news, product reviews and social media, has increased the need for automatic methods to analyze it quickly and accurately. This analysis is crucial for decision making, so great efforts are devoted to identifying the sentiment of opinions in these unstructured texts. However, in the current context, where opinions are based on specific aspects of a product or service, it has become essential to go beyond general analysis and analyze the sentiments associated with each individual aspect.

Aspect-based sentiment analysis focuses on understanding user sentiment in textual opinions about specific features of an entity. However, identifying aspects and their polarity is a challenge, especially in the Spanish language. Most of the existing solutions for this task have been developed and evaluated mainly in English, which presents lower accuracy for Spanish.

This thesis proposes a new approach to aspect-based sentiment analysis, using ensembles of Transformers. The central idea is to combine the power of Transformers with the model ensemble technique, in order to take advantage of different architectures and learning strategies. This method consists of three fundamental stages: text preprocessing, aspect extraction and polarity prediction. The solution was evaluated using the SemEval 2016 Task 5 dataset, specifically designed for aspect-based sentiment analysis.

**Keywords:** Deep learning, Transformers, sentiment analysis, aspect extraction, ensemble.

**Anexo A:**

**ANEXO A. PLAN INDIVIDUAL DE TAREAS DEL ESTUDIANTE**

|  |  |  |
| --- | --- | --- |
| **Tareas** | **Fecha de entrega** | **Rol(es) que desarrolla(n) con la tarea** |
| Reunión de inicio de la práctica | 29/04/24 | - |
| Actualizar la caracterización de los trabajos publicados relacionados con el análisis de sentimiento basado en aspectos, con énfasis en aquellos que usan arquitectura de Transformer | 10/05/24 | AN |
| Implementar alternativas de solución de análisis de sentimiento basado en aspectos usando diferentes modelos de Transformers, considerando una misma arquitectura para las dos tareas. | 07/06/24 | PG |
| Evaluar las diferentes alternativas implementadas con varias colecciones de prueba y comparar los resultados con lo reportado para esas colecciones de prueba. | 14/06/24 | PB |
| Corregir errores y refinar soluciones desarrolladas | 21/05/24 | PG |
| Elaborar informe de la práctica | 05/07/24 | EE |
| Entregar informe de la práctica al tutor | 08/07/24 | EE |
| Rectificar señalamientos del informe | 19/07/24 | EE |
| Entrega del informe final de la práctica | 09/09/24 | EE |
| Defensa de la práctica | 14-18/09/24 | Todos |

|  |  |  |
| --- | --- | --- |
| Alfredo Simón Cuevas  Nombre completo y firma del primer tutor | Miguel Ángel Rivero Tapia  Nombre completo y firma del estudiante |  |

|  |
| --- |
| En la columna Rol se deben poner la sigla del rol que contribuye a formar cada tarea. Las siglas de los roles son:  **AN**- Analista de negocio **AR**- Arquitecto **JP**-Jefe de proyecto  **AS**- Analista de sistema **DS**- Diseñador de software **ES**- Especialista de seguridad  **AD**- Analista de datos **DI**- Diseñador de Interfaz hombre – máquina **EE**- Escritor-expositor de trabajos técnicos  **PG**- Programador **DB**- Diseñador de base de datos **II**- Habilitador de Infraestructuras Informáticas  **PB**- Probador **GC**- Gestor de Configuración **TD**- Facilitador de la Toma de Decisiones |

**Índice:**

[Introducción 1](#_Toc176789127)

[Capítulo 1: Análisis de sentimientos basado en aspectos. 8](#_Toc176789128)

[1.1. Procesamiento del Lenguaje Natural (PLN). 8](#_Toc176789129)

[1.2. Análisis de Sentimientos. 9](#_Toc176789130)

[1.2.1. Niveles del análisis de sentimientos. 9](#_Toc176789131)

[1.3. Análisis de sentimientos basado en aspectos. 10](#_Toc176789132)

[1.4. Enfoques utilizados en el análisis de sentimientos basados en aspectos. 11](#_Toc176789133)

[1.4.1. Enfoque no supervisado. 12](#_Toc176789134)

[1.4.2. Enfoque supervisado. 13](#_Toc176789135)

[1.5. Trabajos relacionados. 15](#_Toc176789136)

[1.6. Redes neuronales Transformers. 16](#_Toc176789137)

[1.7. Modelos pre-entrenados de lenguaje con arquitectura Transformers. 18](#_Toc176789138)

[1.7.1. BERT (*Bidirectional Encoder Representation from Transformers*). 19](#_Toc176789139)

[1.7.2. ALBERT (*A Lite BERT*). 20](#_Toc176789140)

[1.7.3. RoBERTa (*Robustly optimized BERT approach*). 20](#_Toc176789141)

[1.7.4. GPT-2 (*Generative Pre-trained Transformer 2*). 21](#_Toc176789142)

[1.7.5. ELECTA (Efficiently Learning an Encoder that Classifies Token Replacements Accurately). 22](#_Toc176789143)

[1.8. Técnicas de ensamble de clasificadores. 23](#_Toc176789144)

[1.8.1. Método de votación (*Voting Ensemble*). 23](#_Toc176789145)

[1.8.1.1. Votación máxima. 24](#_Toc176789146)

[1.8.1.2. Votación promedio. 24](#_Toc176789147)

[1.8.1.3. Votación promedio ponderada. 25](#_Toc176789148)

[1.8.2. Método de metaaprendizaje. 26](#_Toc176789149)

[1.8.3. Método de embolsado (*Bagging*). 26](#_Toc176789150)

[1.8.4. Método de apilamiento (*Stacking*). 27](#_Toc176789151)

[Métodos de evaluación. 27](#_Toc176789152)

[1.8.5. Dataset. 28](#_Toc176789153)

[1.8.6. Métricas utilizadas 29](#_Toc176789154)

[1.8.6.1. Precisión 29](#_Toc176789155)

[1.8.6.2. Recall 30](#_Toc176789156)

[1.8.6.3. F1-Score 30](#_Toc176789157)

[1.8.6.4. Accuracy 30](#_Toc176789158)

[1.9. Conclusiones parciales. 31](#_Toc176789159)

[Capítulo 2: Solución de análisis de sentimientos basado en aspectos para idioma español. 32](#_Toc176789160)

[2.1. Descripción de la solución. 32](#_Toc176789161)

[2.1.1. Preprocesamiento 33](#_Toc176789162)

[2.1.2. Extracción de aspectos. 34](#_Toc176789163)

[2.2. Desarrollo de la solución de análisis de sentimientos basado en aspectos. 35](#_Toc176789164)

[2.2.1. Modelo de dominio. 35](#_Toc176789165)

[2.2.2. Glosario de términos. 36](#_Toc176789166)

[2.2.3. Captura de requisitos. 36](#_Toc176789167)

[2.2.3.1. Requisitos funcionales. 36](#_Toc176789168)

[2.2.3.2. Requisitos no funcionales. 37](#_Toc176789169)

[2.2.4. Diagrama de caso de uso del sistema. 38](#_Toc176789170)

[2.2.5. Descripción detallada de los casos de uso. 39](#_Toc176789171)

[2.2.6. Descripción de paquetes. 45](#_Toc176789172)

[2.2.7. Descripción de clases. 46](#_Toc176789173)

[2.3. Problemas frecuentes. 47](#_Toc176789174)

[2.3.1. Insuficiencia de datos. 47](#_Toc176789175)

[2.3.2. Desbalance de clases. 47](#_Toc176789176)

[2.3.3. Ruido en los datos. 48](#_Toc176789177)

[2.3.4. Sobreajuste (*Overfitting*). 48](#_Toc176789178)

[2.4. Conclusiones parciales. 48](#_Toc176789179)

[Capítulo 3: Evaluación de la solución propuesta. 49](#_Toc176789180)

[3.1. Marco de Evaluación. 49](#_Toc176789181)

[3.2. Resultados y discurso. 50](#_Toc176789182)

[3.2.1. Resultados de la evaluación de la extracción de aspectos. 50](#_Toc176789183)

[3.2.2. Análisis de los resultados de la extracción de aspectos. 51](#_Toc176789184)

[3.3. Conclusiones parciales. 53](#_Toc176789185)

[Conclusiones generales. 53](#_Toc176789186)

[Recomendaciones. 55](#_Toc176789187)

[Referencias Bibliográficas 56](#_Toc176789188)

**Índice de tablas.**

[Tabla 1. Soluciones de ABSA reportadas en la literatura 15](#_Toc176769772)

[Tabla 2. Glosario de termino. 36](#_Toc176769773)

[Tabla 3. Descripción del CUS “Seleccionar fuente de datos”. 39](#_Toc176769774)

[Tabla 4. Descripción del CUS “Preprocesar datos”. 39](#_Toc176769775)

[Tabla 5. Descripción del CUS “Entrenar modelos”. 40](#_Toc176769776)

[Tabla 6. Descripción del CUS “Modelar para aspectos”. 40](#_Toc176769777)

[Tabla 7. Descripción del CUS “Modelar para clasificación”. 41](#_Toc176769778)

[Tabla 8. Descripción del CUS “Predecir polaridad”. 42](#_Toc176769779)

[Tabla 9. Descripción del CUS “Extraer aspectos”. 43](#_Toc176769780)

[Tabla 10. Descripción del CUS “Clasificar opinión”. 44](#_Toc176769781)

[Tabla 11.Descripción del CUS “Visualizar resultados”. 44](#_Toc176769782)

[Tabla 12. Descripcion de paquetes 45](#_Toc176769783)

[Tabla 13. Descripcion de las clases utilizadas. 46](#_Toc176769784)

[Tabla 14. Configuracion de los hiperparametros de los modelos. 50](#_Toc176769785)

[Tabla 15. Evaluación de la calidad de la extracción de aspectos en el dataset de SemEval2016 51](#_Toc176769786)

[Tabla 16. Comparación de la solución para la extracción de aspectos frente a otras soluciones reportadas en la literatura. 52](#_Toc176769787)

**Índice de figuras.**

[Figura 1. Flujo de ensamble por votación máxima. 24](#_Toc176772764)

[Figura 2. Esquema general del flujo del ABSA en opiniones utilizando un enfoque supervisado. 32](#_Toc176772765)

[Figura 3. Representacion de las reseñas. 34](#_Toc176772766)

[Figura 4. Modelo de dominio. 36](#_Toc176772767)

[Figura 5.Diagrama de casos de uso. 39](#_Toc176772768)

# Introducción

La proliferación de datos textuales en internet tales como blogs, prensa en línea, webs comparativas, evaluación de productos y servicios, especialmente en redes sociales, ha generado un interés creciente en el análisis de sentimientos. Este interés se debe a que la información que se genera en línea, a través de textos, refleja las opiniones, emociones y valoraciones de las personas sobre un sinfín de temas. Esta información se convierte en un recurso valioso para diversas áreas, como el marketing, la gestión de la reputación, la investigación de mercados y la atención al cliente.

El análisis de sentimientos, es el campo de estudio que analiza las opiniones, sentimientos, valoraciones, actitudes y emociones de las personas hacia entidades como productos, servicios, organizaciones, individuos, problemas, eventos, temas y sus atributos. En la literatura se le conoce con muchos nombres y tareas ligeramente diferentes, por ejemplo, minería de opiniones, extracción de opiniones, minería de sentimientos, análisis de subjetividad, análisis de afectos, análisis de emociones, minería de reseñas, entre otras [[1](#_ENREF_1), [2](#_ENREF_2)]

Su relevancia se extiende a diversas áreas, como el análisis de reseñas de productos, la monitorización de la opinión pública o la predicción de las tendencias de consumo. El análisis de sentimientos proporciona a empresas y organizaciones una herramienta poderosa para comprender mejor la percepción del público y tomar decisiones estratégicas más acertadas[[2](#_ENREF_2), [3](#_ENREF_3)].

El análisis de sentimientos ha sido objeto de estudio desde diversas perspectivas, centrándose principalmente en tres niveles: documento, oración y aspecto. Cada uno de estos niveles ofrece un grado de detalle diferente en el análisis de las opiniones expresadas en los textos, permitiendo una comprensión más profunda de la información emocional que subyace en el lenguaje [[2](#_ENREF_2), [4](#_ENREF_4)].

En el nivel de documento, el objetivo principal es determinar la polaridad general de un texto. Se busca clasificar si el documento, en su conjunto, expresa un sentimiento positivo o negativo [[3](#_ENREF_3), [5](#_ENREF_5)].

El análisis de sentimientos a nivel de oración profundiza en el análisis de textos, examinando cada frase individualmente. Su objetivo es determinar si cada oración expresa un sentimiento positivo, negativo o neutral [[3](#_ENREF_3), [5](#_ENREF_5)].

Aunque los niveles de documento y oración proporcionan información valiosa, no permiten identificar con precisión qué aspectos específicos generan satisfacción o insatisfacción en los usuarios. En lugar de analizar estructuras lingüísticas como documentos, párrafos u oraciones, este nivel se centra en la propia opinión, identificando el sentimiento y el objeto al que se dirige [[3](#_ENREF_3), [5](#_ENREF_5)].

El creciente interés en el Análisis de Sentimiento Basado en Aspectos (por sus siglas en ingles *ABSA*) en los últimos años ha impulsado la investigación en este campo, especialmente en el idioma inglés. Sin embargo, otros idiomas como el español, chino o checo, entre otros, aún no han recibido la misma atención. El *ABSA* busca extraer los términos de aspectos de una entidad y determinar la polaridad del sentimiento asociado a dicha entidad. Un aspecto se refiere a un atributo o componente específico de la entidad. Por ejemplo, en la frase "La carne de este restaurante es exquisita", el aspecto sería "carne", la entidad sería "restaurante" y el sentimiento asociado sería "exquisito", con una polaridad positiva [[6](#_ENREF_6)].

La extracción de términos de aspecto (por sus siglas en ingles *ATE*) es una subtarea fundamental en el análisis de sentimientos basado en aspectos. Su objetivo principal radica en identificar y extraer los aspectos relevantes presentes en un texto, sin contar con una lista predefinida de dichos aspectos, teniendo en cuenta que los aspectos pueden ser explícitos o implícitos. Esta tarea constituye la etapa inicial del proceso de *ABSA* y se considera una de las más desafiantes debido a su complejidad [[7](#_ENREF_7), [8](#_ENREF_8)].

La *ATE* busca descubrir y capturar los elementos clave que representan los aspectos mencionados en el texto. Estos aspectos pueden variar ampliamente dependiendo del dominio o contexto en el que se aplique. Al realizar una extracción precisa de los términos de aspecto, se facilita enormemente la comprensión y la clasificación del sentimiento expresado en el texto [[7](#_ENREF_7), [8](#_ENREF_8)].

El campo del Procesamiento del Lenguaje Natural (PLN) ha experimentado un avance sin precedentes en las últimas décadas, impulsado en gran medida por la aplicación de métodos de aprendizaje automático. Esta transformación ha permitido desarrollar sistemas capaces de comprender, interpretar y generar lenguaje humano con una precisión y sofisticación nunca antes vistas. Tradicionalmente, el PLN se basaba en reglas gramaticales y análisis lingüísticos, lo que limitaba su capacidad para manejar la complejidad y la ambigüedad del lenguaje natural. Sin embargo, la llegada del aprendizaje automático ha revolucionado la forma en que se aborda el PLN [[9](#_ENREF_9)].

El aprendizaje automático, también conocido como *Machine Learning* (*ML*), ha permitido superar estas limitaciones al permitir que los sistemas aprendan de manera autónoma a partir de grandes conjuntos de datos. Los algoritmos de *ML*, como la regresión lineal, los árboles de decisión y las máquinas de vectores de soporte, han demostrado ser efectivos para tareas como la clasificación de textos, la detección de spam y la extracción de información [[9](#_ENREF_9)].

Sin embargo, el desarrollo del aprendizaje profundo (*Deep Learning*, *DL*) ha impulsado aún más el avance del PLN. Los algoritmos de *DL*, basados en redes neuronales, han demostrado una capacidad excepcional para aprender patrones complejos a partir de grandes cantidades de datos, permitiendo la creación de modelos de lenguaje más precisos y versátiles. Esta capacidad ha sido fundamental para el desarrollo de aplicaciones de *PLN* de vanguardia, como los sistemas de traducción automática, los asistentes virtuales, los *chatbots* de servicio al cliente y los sistemas de análisis de sentimientos [[10](#_ENREF_10), [11](#_ENREF_11)] .

La popularidad del aprendizaje profundo, en particular, ha dirigido la atención hacia arquitecturas de redes neuronales basadas en capas convolucionales y recurrentes para la comprensión y procesamiento del lenguaje. Estas arquitecturas permitieron representaciones más ricas y detalladas en comparación con los modelos tradicionales de *Machine Learning*. Las redes convolucionales, por ejemplo, facilitan el análisis del texto por partes, mediante filtros, mientras que las redes recurrentes procesan el texto secuencialmente, considerando las salidas de partes anteriores como entrada adicional [[11](#_ENREF_11)].

Sin embargo, la aparición de los modelos de redes neuronales basados en la atención, ha reorientado la investigación en aprendizaje profundo para PLN hacia el desarrollo de estructuras transformadoras [[11](#_ENREF_11)].

Estas redes neuronales llamadas *Transformers* son un tipo de modelo de aprendizaje profundo que se utilizan principalmente en el procesamiento del lenguaje natural. Su arquitectura fue propuesta en el trabajo "*Attention Is All You Need*" (2017) [[12](#_ENREF_12)]. La principal característica de las redes neuronales *Transformers* es que utilizan un mecanismo llamado atención para procesar secuencias de datos de longitud variable, como, por ejemplo, frases o párrafos. En lugar de procesar las palabras de forma secuencial, como se hace en los modelos basados en redes neuronales recurrentes (RNN), las redes neuronales *Transformers* pueden procesar todas las palabras en paralelo. Esta capacidad de procesamiento paralelo, junto con el mecanismo de atención, ha llevado a los *Transformers* a alcanzar niveles de precisión y eficiencia nunca antes vistos en el PLN [[11](#_ENREF_11), [12](#_ENREF_12)].

Estos modelos, pre-entrenados con millones de parámetros, requieren solo un ajuste fino para adaptarse a una amplia gama de tareas de *PLN*. Además, requieren menos datos y tiempo de entrenamiento que otros modelos, lo que facilita su implementación y escalabilidad. Su impacto en el *PLN* es notable: son la base de muchos modelos de lenguaje de última generación, como *GPT-4, BERT* y *XLNet*, que impulsan herramientas como *GitHub Copilot*, y la última generación de motores de búsqueda de Google [[10](#_ENREF_10), [11](#_ENREF_11)].

Para mejorar el rendimiento de estos modelos pre-entrenados se estará utilizando una técnica llamada: aprendizaje conjunto, en ingles *Ensemble Learning*, esta es una poderosa estrategia en el aprendizaje automático. Se basa en integrar múltiples modelos de referencia dentro de un mismo marco con el objetivo de crear un modelo "maestro" que supere en rendimiento a cada uno de los modelos individuales. Al combinar las fortalezas de varios modelos, se obtiene un sistema más robusto y preciso. Este enfoque se ha convertido en una tendencia fundamental en la investigación del aprendizaje automático, impulsado por su capacidad de lograr resultados de vanguardia en diversas aplicaciones. El éxito de un método de aprendizaje conjunto depende de dos aspectos clave: el entrenamiento de los modelos base y la combinación de las predicciones [[13](#_ENREF_13), [14](#_ENREF_14)].

Tras una exhaustiva revisión de la literatura, se han detectado dos patrones relevantes en el ámbito del análisis de sentimientos basado en aspectos: primero, la subtarea de extracción de aspectos presenta un desarrollo significativamente menor en comparación con la clasificación de polaridad, con menos investigaciones dedicadas a ella, lo que se traduce en niveles de precisión más bajos en general. Este desequilibrio se acentúa en el contexto del idioma español, donde la escasez de estudios y la consecuente reducción en los niveles de precisión son aún más pronunciadas. Segundo, las técnicas de aprendizaje conjunto han demostrado su eficacia en la mejora del rendimiento de los modelos de análisis de sentimientos. Sin embargo, la aplicación de estas técnicas a la extracción de aspectos en español presenta una carencia notable. No se han encontrado estudios que exploren esta combinación para textos en español, a diferencia de la clasificación de polaridad, donde existen investigaciones que han implementado el aprendizaje conjunto con resultados positivos.

A partir de la situación anterior se plantea como **problema** a resolver la existencia de una baja precisión en la extracción de aspectos para el análisis de sentimientos basado en aspectos en textos en español, especialmente en comparación con la clasificación de polaridad.

El **objeto de estudio** de la presente investigación se enmarca en el análisis de sentimientos. Se identifica como **campo de acción** el análisis de sentimientos basado en aspectos con el uso de modelos *Transformers* y técnicas de aprendizaje conjunto.

El **objetivo general** de esta investigación consiste en desarrollar un algoritmo de análisis de sentimientos basado en aspectos para el idioma español con el uso de técnicas de aprendizaje conjunto en modelos *Transformers* que mejore la precisión en la extracción de aspectos en comparación con los trabajos reportados en la literatura.

Este se desglosa en los siguientes **objetivos específicos:**

1. Caracterizar los fundamentos teóricos y los trabajos relacionados al análisis de sentimientos basado en aspectos.

* Tareas:

1. Estudiar los fundamentos teóricos del análisis de sentimientos que utilicen un enfoque basado en aspectos.
2. Estudiar los fundamentos teóricos de la técnica de aprendizaje conjunto.
3. Identificar trabajos previos relacionados con la extracción de aspectos y clasificación de la polaridad.
4. Seleccionar y estudiar los modelos de *Transformers* para la extracción de aspectos.
5. Desarrollar una nueva solución de análisis de sentimientos basado en aspectos con el uso de modelos *Transformers* y técnicas de ensamble que permita mejorar los resultados de soluciones propuestas en la literatura para idioma español.

* Tareas:

1. Aplicar la limpieza de los textos a utilizar para el aprendizaje del modelo.
2. Concebir un mecanismo de extracción de aspectos.
3. Implementar el método seleccionado.
4. Diseñar una solución para la clasificación de la polaridad.
5. Entrenar el modelo con los datos de entrenamiento
6. Evaluar experimentalmente la solución utilizando colecciones de prueba reconocidos.

* Tarea:

1. Seleccionar las métricas a utilizar para medir la efectividad del modelo.
2. Selección los datos para realizar los experimentos.
3. Efectuar experimentos con el nuevo modelo utilizando las métricas seleccionadas y los dataset.
4. Comparar los resultados obtenidos del modelo con otros modelos existentes.

El informe está estructurado en tres capítulos:

**Capítulo 1** – Análisis de Sentimientos Basado en Aspectos: Se describen de manera general los conceptos del Procesamiento del Lenguaje Natural y el Análisis de Sentimientos Basado en Aspectos. Se muestran los enfoques más utilizados para la extracción de aspectos y clasificación de la polaridad. Se describen los modelos *Transformers* y se muestran los trabajos relacionados.

**Capítulo 2** - Solución de análisis de sentimientos basado en aspectos para idioma español: En este capítulo se realiza una descripción detalla de cada uno de los procedimientos realizados para dar solución a la problemática planteada. Se dan a conocer cuáles fueron los modelos *Transformers* utilizados. Además, se muestra la documentación asociada a la solución mediante la utilización de diferentes artefactos de la ingeniería de software.

**Capítulo 3** - Evaluación de la solución propuesta: Se muestra cuál fue el marco de evaluación utilizado. Se describen cada una de las posibles configuraciones a realizar para la evaluación. Se comparan los resultados obtenidos en cada uno de los algoritmos propuestos utilizando las métricas de evaluación y se muestra cuál fue el resultado del método propuesto.

# Capítulo 1: Análisis de sentimientos basado en aspectos.

En este capítulo se explican los conceptos teóricos más importantes del Procesamiento del Lenguaje Natural y el Análisis de Sentimientos Basado en Aspectos, profundizando en la problemática de la Extracción de Aspectos. Se caracterizan soluciones afines a la extracción de aspectos enfocadas en el aprendizaje profundo, reportadas en la literatura. Se realiza una introducción a las principales técnicas y métodos empleados, así como sus principales limitaciones.

## Procesamiento del Lenguaje Natural (PLN).

El procesamiento del lenguaje natural (, denominado también *NLP* por sus siglas en inglés) es un área de investigación y aplicación de la Minería de Textos que explora cómo las computadoras se pueden usar para comprender y manipular texto del lenguaje natural. Se puede entender como la habilidad de la máquina para procesar la información comunicada, no simplemente las letras o los sonidos del lenguaje [[15](#_ENREF_15), [16](#_ENREF_16)].

Es una gama de técnicas computacionales teóricamente motivadas para analizar y representar textos que ocurren naturalmente en uno o más niveles de análisis lingüístico con el fin de lograr un procesamiento del lenguaje similar al humano para una variedad de tareas o aplicaciones. Los cimientos del *PLN* se encuentran en una serie de disciplinas como son: la informática y ciencias de la información, lingüística, matemáticas, ingeniería eléctrica y electrónica, inteligencia artificial y robótica, psicología, entre otras [[15](#_ENREF_15), [16](#_ENREF_16)].

El procesamiento del lenguaje natural proporciona teoría e implementaciones para una gran variedad de aplicaciones de la vida real. Entre las más frecuentes se incluyen las siguientes [[15](#_ENREF_15)]:

* Extracción de información.
* Sistemas de diálogo.
* Análisis de Sentimientos.

## Análisis de Sentimientos.

El análisis de sentimiento (por sus siglas en ingles, *SA*) o minería de opinión es el estudio computacional de opiniones, sentimientos y emociones expresados en texto. Como campo de investigación, está estrechamente relacionada con la lingüística computacional, procesamiento del lenguaje natural y la minería de textos. Se encarga de explorar y descubrir la información subjetiva generada por el usuario. Como campo de estudio, aglomera un conjunto de técnicas para recopilar, agrupar, estructurar, analizar y valorar texto a fin de conocer cómo el ser humano utiliza el lenguaje para expresar información afectiva [[17](#_ENREF_17), [18](#_ENREF_18)].

La investigación en el análisis de sentimientos cubre casi todas las áreas de investigación del procesamiento del lenguaje natural, cubre semántica léxica, análisis semántico, extracción de información, análisis del discurso, entre otras y tiene como objetivo fundamental encontrar las opiniones, identificar sus sentimientos y clasificarlos en opiniones positivas y negativas [[19](#_ENREF_19)].

### Niveles del análisis de sentimientos.

El componente principal del *SA* es el proceso de clasificación de opiniones en función de su polaridad. La clasificación se puede realizar en diferentes niveles, incluido el nivel del documento, el nivel de la oración o frase y el nivel del aspecto, en el que cada nivel se considera como la unidad de información para la tarea de clasificación.

* **Nivel de documento.**

Se realiza en un documento completo y se otorga una polaridad única a todo el documento. Se clasifican en positivo, negativo o neutro documentos como: páginas de libros, blogs, críticas de productos entre otros. Pueden usarse enfoques de aprendizaje automático (supervisados, no supervisados) para clasificar el documento. La clasificación a nivel de documento funciona mejor cuando el documento está escrito por una sola persona y no es adecuada cuando evalúan o comparan varias entidades [[17](#_ENREF_17), [18](#_ENREF_18)].

* **Nivel de Oración.**

A nivel de oración, cada una de las oraciones es analizada y asignada una polaridad correspondiente, esta polaridad se determinará de forma independiente utilizando las mismas metodologías que el nivel de documento, pero con mayores datos de entrenamiento y recursos de procesamiento, pudiendo utilizarse para encontrar el sentimiento del documento o usarse individualmente. Es muy útil cuando un documento tiene una amplia gama y mezcla de sentimientos asociados. Este nivel de clasificación está asociado con la clasificación subjetiva. Es fundamental en trabajos donde se quiere encontrar oraciones subjetivas, oraciones condicionales o declaraciones ambiguas [[3](#_ENREF_3), [5](#_ENREF_5)].

* **Nivel de Aspecto.**

El análisis de sentimiento se realiza a nivel de aspecto cuando cada oración puede contener múltiples aspectos, se les asigna polaridad a todos los aspectos por independiente y después se calcula el sentimiento agregado para toda la oración. Este se lleva a cabo con mayor granularidad; ya que extrae los atributos de la entidad de la cual se expresa una opinión y la orientación de dicha opinión [[3](#_ENREF_3), [5](#_ENREF_5)].

## Análisis de sentimientos basado en aspectos.

El Análisis de Sentimiento Basado en Aspectos se erige como una herramienta sofisticada dentro del campo del análisis de sentimiento, ofreciendo una comprensión más profunda y granular de la opinión expresada en un texto. A diferencia de la simple clasificación de sentimientos positivos o negativos, el *ABSA* se centra en analizar la polaridad del sentimiento hacia características específicas, o aspectos, de un producto, servicio, persona o evento. Un aspecto se define como un atributo relevante que contribuye a la evaluación de un objeto. Por ejemplo, en una reseña de un restaurante, los aspectos podrían ser la comida, el servicio, el ambiente o el precio [[20](#_ENREF_20), [21](#_ENREF_21)]

En general el análisis de sentimientos basado en aspectos está formado por dos tareas principales [[22](#_ENREF_22)].

* **Extracción de Aspectos:** Esta etapa consiste en identificar las características o atributos específicos, llamados "aspectos", que se mencionan en un texto. Por ejemplo, en una reseña de un restaurante, los aspectos podrían ser la calidad de la comida, el servicio, la decoración, el precio, entre otros. La extracción de aspectos es un proceso desafiante, ya que requiere la capacidad de comprender el contexto del texto y reconocer los elementos relevantes que se están evaluando.
* **Clasificación de Sentimientos:** Una vez que se han identificado los aspectos, se debe determinar la polaridad del sentimiento expresado hacia cada uno de ellos. En otras palabras, se busca determinar si la opinión sobre el aspecto es positiva, negativa o neutral. Esta tarea requiere analizar el contexto lingüístico alrededor del aspecto para inferir la intención del autor.

## Enfoques utilizados en el análisis de sentimientos basados en aspectos.

El análisis de sentimientos basado en aspectos ha experimentado un auge en los últimos años, impulsado por la necesidad de comprender las opiniones de los usuarios con mayor profundidad. La investigación en este campo ha dado lugar a una variedad de enfoques, cada uno con sus propias fortalezas y limitaciones. Algunos de los métodos más utilizados incluyen modelos basados en redes neuronales, técnicas de aprendizaje profundo y los modelos de *Transformers*. Estos modelos, a menudo entrenados con datos etiquetados, pueden tener un enfoque supervisado o no supervisado. La elección del enfoque depende de la disponibilidad de datos, la complejidad de la tarea y los objetivos específicos de la investigación [[23](#_ENREF_23)].

### Enfoque no supervisado.

El ABSA no supervisado se caracteriza por su capacidad de extraer aspectos y clasificar la polaridad del sentimiento sin depender de conjuntos de datos pre-etiquetados. Estas técnicas se basan en la identificación de patrones y relaciones dentro del texto para inferir la estructura de la opinión [[21](#_ENREF_21)].

La investigación pionera de [[24](#_ENREF_24)] sentó las bases para un enfoque no supervisado para el resumen de sentimientos basado en aspectos, donde la detección de aspectos se considera una tarea fundamental. Su método se basa en la identificación de aspectos mediante el análisis de patrones basados en sustantivos, lo que permite la extracción de información relevante sobre los diferentes aspectos de un producto o servicio. Esta investigación fue fundamental para el desarrollo de técnicas no supervisadas para el análisis de sentimientos, abriendo camino a nuevas investigaciones en la extracción de aspectos y el resumen de opiniones.

En [[25](#_ENREF_25)] se propone un enfoque basado en ontologías para la extracción de aspectos y utiliza el recurso lingüístico *SentiWordNet* para la clasificación de polaridad. Este enfoque se destaca por su capacidad de considerar las palabras negativas en el análisis, lo que se traduce en una mayor precisión en la detección de sentimientos. La incorporación de ontologías y *SentiWordNet* permite una comprensión más profunda del contexto y las relaciones semánticas entre las palabras, lo que resulta en una mejor interpretación de la polaridad de los comentarios. Este enfoque representa un avance en la precisión del análisis de sentimientos al abordar la complejidad de las expresiones negativas, que a menudo son difíciles de procesar por los métodos tradicionales.

Por otro lado [[26](#_ENREF_26)] presenta un enfoque prometedor para la extracción de aspectos mediante la combinación de *Latent Dirichlet Allocation* con un modelo pre-entrenado no supervisado. Este método destaca por su independencia lingüística y contextual, atributo poco común en el campo del análisis de sentimientos. Sin embargo, el sistema propuesto presenta limitaciones. Requiere una intervención manual para proporcionar información sobre palabras relacionadas con el dominio específico bajo análisis, lo que introduce un riesgo de error humano y afecta la consistencia de los resultados. La dependencia de la entrada manual en este sistema limita su capacidad de generalización y robustez. Para mitigar este problema, futuras investigaciones podrían explorar alternativas que exploren la capacidad de aprendizaje automático para la detección automática de términos relevantes al dominio, reduciendo así la necesidad de intervención humana y aumentando la precisión y consistencia del análisis.

### Enfoque supervisado.

El aprendizaje supervisado se basa en la construcción de una relación entre variables de entrada (X) y una variable de salida (Y). Esta relación se aprende a partir de datos etiquetados, donde cada ejemplo está asociado a una etiqueta que indica la clase o valor deseado. El objetivo es utilizar esta relación aprendida para predecir la salida de nuevos datos no vistos [[27](#_ENREF_27)].

La esencia del aprendizaje supervisado reside en la capacidad de generar predicciones precisas sobre datos futuros basándose en el análisis de información previamente etiquetada. Este proceso implica una fase de entrenamiento, donde los algoritmos aprenden la relación entre las variables, y una fase de prueba, donde se evalúa la precisión de las predicciones sobre datos nuevos [[27](#_ENREF_27), [28](#_ENREF_28)].

Si bien un gran número de trabajos de investigación se han centrado en el desarrollo y aplicación de algoritmos supervisados tradicionales de *Machine Learning* para el análisis de sentimientos basado en aspectos, este epígrafe se enfoca en la exploración de los trabajos que utilizan las redes neuronales profundas, ya que estas son una herramienta más sofisticada para esta tarea.

Las arquitecturas de redes neuronales profundas, como las redes neuronales recurrentes (RNN), las redes neuronales convolucionales (CNN) y las redes neuronales Transformers, han demostrado un gran éxito en la tarea de análisis de sentimientos [[29](#_ENREF_29)].

Por ejemplo, en [[20](#_ENREF_20)] se propone un nuevo modelo para la extracción de aspectos basado en el aprendizaje profundo y permanente, utilizando una técnica de reducción catastrófica del olvido en múltiples dominios. La arquitectura se basa en una CNN y un modelo de *Lifelong Learning*, compuesto por cuatro etapas. En la primera etapa, las opiniones textuales se convierten en vectores *Word Embedding*. La segunda etapa, la extracción de conocimientos básicos, entrena la CNN para cada dominio y actualiza la Base de Conocimiento (KB). La etapa tres, la actualización de la KB, utiliza un análisis de pérdidas para evitar el olvido catastrófico, enriqueciendo la KB con nuevos aspectos extraídos de los diferentes dominios. Finalmente, la etapa cuatro crea el extractor de aspectos, un modelo CNN multidominio que se configura con los parámetros comunes de la KB.

Del mismo modo, [[21](#_ENREF_21)] propone un modelo llamado BERT *Adversarial* Training (BAT) para mejorar la extracción de aspectos y la clasificación de sentimientos de aspectos en análisis de sentimientos. El modelo utiliza un entrenamiento adversario para ajustar el BERT de propósito general y el BERT post-entrenado específico de dominio (BERT-PT). El entrenamiento adversario implica crear ejemplos contradictorios a partir de las incrustaciones de BERT utilizando el gradiente de pérdida, y luego alimentar estos ejemplos perturbados al codificador BERT para calcular la pérdida adversaria. Finalmente, se aplica el algoritmo de retropropagación a la suma de la pérdida de entrenamiento y la pérdida adversaria. Los experimentos realizados muestran que BAT supera al BERT general y al BERT post-entrenado en ambas tareas, tanto en el ámbito de restaurantes como de portátiles.

También en [[30](#_ENREF_30)] se propone un marco de aprendizaje conjunto basado en BERT para abordar todas las subtareas del ABSA. En lugar de tratar cada subtarea por separado, este marco utiliza dos modelos BERT-MRC (*Machine Reading Comprehension*) con parámetros compartidos para resolver las tareas de extracción de aspectos y clasificación de polaridad de sentimiento. Los modelos se entrenan conjuntamente para aprender a identificar las posiciones inicial y final de los aspectos y opiniones en una oración, así como la polaridad del sentimiento expresado.

## Trabajos relacionados.

Este epígrafe presenta en la Tabla 1 una revisión exhaustiva de los trabajos relacionados en el ámbito del análisis de sentimientos basado en aspectos utilizando modelos de lenguaje de aprendizaje profundo.

Tabla 1. Soluciones de ABSA reportadas en la literatura

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Referencia** | **Idioma** | **Modelos Deep Learning** | **Dataset** | **F1 (%)** | **Accuracy (%)** |
| [[20](#_ENREF_20)] | Ingles | Convolucional CNN | Amazon y tripAdvisor | Restaurante: 62,5  Hotel. 67,9 | Restaurante: 74,8  Hotel: 78,6 |
| [[31](#_ENREF_31)] | Ingles | T5 | ALSC (Rest16 y MAMS). | 79.71 | -- |
| [[32](#_ENREF_32)] | Ingles | Wo-BiLSTM-CRF-Glove.42B | SemeEval-14 | 86,05 | 81,08 |
| [[33](#_ENREF_33)] | Italiano | RoBERTa | Plataforma de comercio electrónico. | 61.87 | -- |
| [[34](#_ENREF_34)] | hindi | Attention-based BiLSTM | Opiniones de artículos electrónicos | 43.16 | -- |
| [[21](#_ENREF_21)] | Ingles | BERT Adversarial Training (BAT) | SemEval 2014 (Task 4) | Laptop: 85.57  Restaurante:81,5 | -- |
| [[35](#_ENREF_35)] | Ingles | BERT post-training | SemEval 2014 (task 4) | Laptop: 84.26  Restaurante: 77,9 | -- |
| [[36](#_ENREF_36)] | Ingles | DomBERT | SemEval 2014 (Task 4) | Rest : 77.21 | -- |
| [[37](#_ENREF_37)] | Ingles | BART | SemEval 2014 - 2015 | Rest 14: 87.07  Rest 15: 75.48  Lapt 14: 83.52 | -- |
| [[38](#_ENREF_38)] | Ingles | BERT-MTL-  GRU | SemEval 2014 | Restaurante::86.19 | -- |
| [[30](#_ENREF_30)] | Ingles | Bert- Dual-MRC | SemEval 2014 - 2015 | Rest 14: 86.60  Rest 15: 75.08  Lapt 14: 82.51 | -- |
| [[39](#_ENREF_39)] | Español | -- | Semeval 2016 | Restaurante: 73,07 | -- |
| [[40](#_ENREF_40)] | Español | Convolucional (CNN) | Semeval 2016 | Restaurante:65,40 | -- |
| [[22](#_ENREF_22)] | Español | BERT | Semeval 2016 | Restaurante: 76,70 | -- |

## Redes neuronales Transformers.

Las redes neuronales *Transformers* han transformado el campo del procesamiento del lenguaje natural. Su arquitectura innovadora, basada en un mecanismo de atención llamado "autoatención" (*self-attention*), les permite procesar secuencias de datos de longitud variable, como frases o párrafos, de manera más eficiente que los modelos tradicionales basados en redes neuronales recurrentes (RNN) [[12](#_ENREF_12)].

En lugar de procesar las palabras secuencialmente, como en los RNN, los *Transformers* pueden procesar todas las palabras en paralelo, lo que los hace mucho más rápidos y efectivos. Esta capacidad de capturar relaciones complejas entre las palabras a través de la atención les permite generar representaciones contextuales más ricas y precisas [[12](#_ENREF_12)].

La arquitectura de un *Transformers* se basa en bloques de atención que se apilan en capas. Cada bloque *Transformers* se compone de dos subcapas principales [[12](#_ENREF_12), [22](#_ENREF_22)]:

1. Capa de Atención: En esta capa, se calculan los pesos de atención para cada palabra en la secuencia, ponderando su importancia en función de su relación con las demás palabras. Este proceso se realiza mediante cálculos de producto escalar y una función *Softmax*, que permite que las palabras más relevantes tengan un mayor peso en la representación final.

2. Capa de Alimentación hacia Adelante (*Feed-forward*): Esta capa aplica una transformación lineal y una función de activación no lineal a la salida de la capa de atención, agregando más complejidad y expresividad a la representación de la secuencia.

El proceso de entrenamiento de un *Transformer* utiliza la técnica de retropropagación del error. Esto permite ajustar los pesos de la red neuronal de manera que la salida generada por el modelo se acerque lo más posible a la salida deseada para la tarea en cuestión [[12](#_ENREF_12), [22](#_ENREF_22)].

Este tipo de redes presenta ventajas significativas con respecto a otros tipos de redes neuronales [[8](#_ENREF_8), [22](#_ENREF_22)].

* Capacidad para procesar secuencias largas de texto: Como se mencionaba anteriormente, a diferencia de las arquitecturas tradicionales de redes neuronales recurrentes, que sufren de limitaciones en el procesamiento de secuencias largas, las redes *Transformer* pueden manejar textos de longitud variable con una mayor eficiencia. Esto se debe a su mecanismo de atención, que permite a la red enfocarse en partes específicas de la secuencia de entrada, en lugar de procesarla secuencialmente. Esta capacidad es crucial para el análisis de sentimientos basado en aspecto, ya que las expresiones de sentimientos y opiniones a menudo se distribuyen a lo largo de textos extensos, y la comprensión de la relación entre palabras y frases distantes es fundamental para interpretar el sentimiento general.
* Captura de relaciones a largo plazo entre palabras: Las redes *Transformer* son capaces de capturar relaciones a largo plazo entre las palabras en una secuencia, lo que les permite comprender mejor el contexto y producir resultados más precisos. Este es un aspecto fundamental para el análisis de sentimientos, ya que las opiniones y emociones a menudo se expresan de manera sutil y a través de conexiones indirectas entre palabras en el texto. Por ejemplo, en una frase como "El hotel era bonito, pero la comida no era buena", la red *Transformer* puede comprender que "bonito" se relaciona con "hotel" y "no era buena" se relaciona con "comida", permitiendo una interpretación precisa del sentimiento expresado.
* Escalabilidad y aprendizaje adaptativo: Las redes *Transformer* son altamente escalables y pueden entrenarse con grandes conjuntos de datos, lo que les permite obtener resultados altamente precisos en tareas complejas de PLN. Al ser expuestas a una amplia gama de datos, las redes *Transformer* pueden aprender a identificar patrones y relaciones sutiles en el lenguaje, mejorando su capacidad para interpretar sentimientos expresados en diferentes contextos.
* Transferencia de aprendizaje: Son especialmente útiles para la transferencia de aprendizaje, es decir, la capacidad de utilizar el conocimiento aprendido en una tarea para mejorar el rendimiento en otra tarea relacionada. Esto se debe a que las redes neuronales *Transformers* aprenden representaciones de palabras altamente generalizadas que pueden aplicarse a una amplia variedad de tareas de PLN.

## Modelos pre-entrenados de lenguaje con arquitectura Transformers.

La arquitectura *Transformers* ha revolucionado el campo del Procesamiento del Lenguaje Natural al integrar las mejores prácticas existentes y añadir innovaciones significativas. La inclusión de la atención multicabeza permite al modelo comprender mejor la estructura gramatical del lenguaje, más allá de la simple semántica de las palabras, lo que lo convierte en un modelo más potente para la resolución de problemas de PLN. La influencia del *Transformer* es tan profunda que la mayoría de los modelos de vanguardia en PLN, incluyendo modelos de lenguaje basados en *Transformers*, se basan en sus principios para lograr resultados sobresalientes en tareas de clasificación, generación y comprensión de texto [[29](#_ENREF_29)].

### BERT (*Bidirectional Encoder Representation from Transformers*).

BERT es un modelo de lenguaje pre-entrenado que se destaca por su capacidad de comprender el significado de las palabras dentro del contexto de una oración o párrafo completo. Esta arquitectura, basada en las redes neuronales *Transformers*, utiliza un mecanismo de atención para capturar la relevancia de las palabras en el texto, permitiendo una interpretación más profunda del lenguaje [[41](#_ENREF_41)].

Para entrenar a BERT, se empleó una gran cantidad de datos textuales y se implementaron dos técnicas de aprendizaje auto-supervisado. La primera técnica consiste en predecir palabras faltantes en una oración a partir del contexto circundante, conocida como "máscara de palabras". La segunda técnica consiste en predecir la oración siguiente en un texto. Estas técnicas de auto-supervisión permiten que BERT aprenda a representar el lenguaje de forma más eficiente y precisa [[41](#_ENREF_41), [42](#_ENREF_42)].

BERT no se limita a ser un simple mecanismo de representación textual; es un modelo de lenguaje robusto que, además de generar representaciones de palabras, puede utilizarse como base para una amplia gama de tareas de procesamiento del lenguaje natural, como la clasificación de texto, la extracción de información, el análisis de sentimientos y muchas otras [[41](#_ENREF_41), [43](#_ENREF_43)].

Para aplicar BERT a tareas para las que no fue específicamente entrenado, se utiliza el aprendizaje por transferencia. Este proceso implica tomar el modelo pre-entrenado de BERT, que ha aprendido a predecir la oración siguiente y a completar palabras faltantes, y ajustar finamente sus pesos para adaptarlo a una tarea específica [[43](#_ENREF_43)].

Este proceso de ajuste fino (*fine-tuning*) permite que BERT alcance un alto nivel de precisión en diversas tareas de procesamiento del lenguaje natural. La incorporación de una capa adicional de neuronas a la salida de BERT puede mejorar aún más su rendimiento, lo que permite obtener resultados de vanguardia sin necesidad de modificaciones significativas en la arquitectura del modelo original [[41](#_ENREF_41)].

### ALBERT (*A Lite BERT*).

El modelo ALBERT es una variante del modelo BERT, se presenta como una alternativa más eficiente para el procesamiento del lenguaje natural. Si bien BERT se destaca por su capacidad para comprender el significado de las palabras en contexto, su tamaño y complejidad computacional limitan su aplicación en algunos escenarios. ALBERT aborda este problema al reducir la cantidad de parámetros y mejorar la eficiencia del entrenamiento. ALBERT se basa en los principios de BERT, pero introduce mejoras que optimizan su arquitectura. Su estructura se compone de las siguientes capas [[44](#_ENREF_44)]:

* Capa ALBERT: Esta capa es la base del modelo y utiliza la misma técnica de atención multicabeza que BERT, pero con un número reducido de parámetros.
* Capa LSTM (*Long Short-Term Memory*): Esta capa se encarga de capturar la información secuencial del texto, lo que permite comprender el contexto de las palabras y frases.
* Capa de Discriminación de Tendencia Emocional: Esta capa final realiza la clasificación de la polaridad emocional del texto, es decir, determina si el sentimiento expresado es positivo, negativo o neutral.

### RoBERTa (*Robustly optimized BERT approach*).

RoBERTa es una versión mejorada del modelo BERT que se basa en un entrenamiento más robusto y exhaustivo para lograr un mejor rendimiento en tareas de procesamiento del lenguaje natural [[45](#_ENREF_45)].

En lugar de entrenar con un solo conjunto de datos como BERT, RoBERTa se alimenta con un conjunto de datos mucho más amplio, incluyendo *CC-news* (76GB), *OpenWebText* (38GB) y *Stories* (31GB), lo que le permite aprender de una mayor diversidad de lenguaje. Además, RoBERTa se entrena durante más tiempo y con un tamaño de lote más grande, lo que permite que el modelo procese más información y converja a una solución óptima más rápidamente. También se eliminó la tarea de predicción de la siguiente oración del entrenamiento, lo que permite al modelo enfocarse en otras tareas de comprensión del lenguaje, como la comprensión contextual y la predicción de palabras faltantes [[45](#_ENREF_45), [46](#_ENREF_46)].

RoBERTa utiliza un patrón de enmascaramiento diferente al de BERT, lo que mejora su capacidad de comprensión del lenguaje. Esto permite al modelo generar representaciones más precisas y versátiles, lo que se traduce en un mejor rendimiento en una variedad de tareas de procesamiento del lenguaje natural. RoBERTa está disponible en versiones multilingües, incluyendo español. Algunos ejemplos de modelos pre-entrenados para español incluyen BERTIN, RoBERTa-base y RoBERTa-large. Aunque todos comparten la misma arquitectura, sus diferencias residen en los conjuntos de datos utilizados para su entrenamiento y los mecanismos de entrenamiento específicos [[45](#_ENREF_45), [46](#_ENREF_46)].

### GPT-2 (*Generative Pre-trained Transformer 2*).

El modelo GPT-2, desarrollado por *OpenAI*, es un modelo de lenguaje basado en Transformers capaz de generar secuencias largas de texto, adaptándose al estilo y contenido de una entrada arbitraria. Además, los modelos GPT-2 pueden realizar diversas tareas de PLN, como la clasificación (que es el enfoque de este trabajo) [[47](#_ENREF_47)].

Con 1.5 billones de parámetros, GPT-2 está entrenado para predecir la siguiente palabra, dados todas las palabras previas dentro de un texto. Su arquitectura implementa una red neuronal profunda, donde las entradas son procesadas por múltiples capas de neuronas que activan las capas subsecuentes, generando la salida del modelo en la capa final. Este modelo de lenguaje utiliza la arquitectura Transformer, beneficiándose del mecanismo de atención, lo que permite al modelo enfocarse selectivamente en las piezas de entrada más relevantes [[29](#_ENREF_29), [47](#_ENREF_47)].

La arquitectura *Transformers* de GPT-2 tiene dos modelos dispuestos linealmente, el codificador y el decodificador. El primero está compuesto por una capa de autoatención que aplica el mecanismo de atención al texto de entrada y un tipo de avance (*feed-forward)* que convierte el resultado de la capa de autoatención en una incrustación de menor longitud. El segundo también está compuesto por estas unidades, pero con una capa adicional llamada atención codificador-decodificador, la cual mapea el resultado de la capa de autoatención del decodificador con las incrustaciones producidas por el codificador. De esta manera, todo el texto de entrada se procesa a la vez, proporcionando una memoria más estructurada para manejar dependencias a largo plazo en el texto, contrariamente a las técnicas previas que involucraban mecanismos recurrentes [[29](#_ENREF_29), [47](#_ENREF_47)].

Aunque tanto BERT como GPT-2 se basan en la arquitectura *Transformer*, GPT-2 consiste en bloques de decodificador apilados, mientras que BERT está construido utilizando los bloques de codificador. Al igual que los modelos de lenguaje tradicionales, GPT-2 recibe los vectores de palabras y proporciona las estimaciones para la probabilidad de la siguiente palabra, donde cada palabra tiene el contexto de los tokens previos. Después de la generación de cada token, se agrega a la secuencia de entradas, convirtiéndose en la entrada del modelo en su siguiente paso. BERT, sin embargo, no es autorregresivo, utilizando todo el contexto circundante a la vez. Además, en lugar de una capa de autoatención como BERT, GPT-2 utiliza una autoatención enmascarada, la cual enmascara los tokens futuros bloqueando información de los tokens a la derecha de la posición que se está calculando [[29](#_ENREF_29), [47](#_ENREF_47)].

### ELECTA (Efficiently Learning an Encoder that Classifies Token Replacements Accurately).

ELECTRA representa un avance significativo en el pre-entrenamiento de modelos de lenguaje, introduciendo una nueva estrategia basada en la teoría de juegos. En lugar de utilizar la función de pérdida tradicional del modelo BERT, que se basa en la predicción de tokens enmascarados (*Masked Language Model - MLM*), ELECTRA propone una función de pérdida adversarial [[22](#_ENREF_22)].

Este enfoque adversarial ofrece varias ventajas sobre los métodos tradicionales. En primer lugar, el entrenamiento se realiza sobre todos los tokens de entrada, no solo sobre los enmascarados. Esto permite que el modelo aprenda de manera más eficiente. En segundo lugar, la competencia entre el generador y el discriminador fomenta la generación de representaciones contextuales más robustas [[48](#_ENREF_48)].

ELECTRA introduce un cambio significativo en la función de pérdida, pero no altera la arquitectura del *Transformer.* Sin embargo, la arquitectura del *Transformer* presenta una redundancia secuencial que puede aumentar el coste del pre-entrenamiento. *Funnel Transformer* aborda este problema mediante una nueva arquitectura que reduce la redundancia. El enfoque de *Funnel Transformer* se basa en la idea de comprimir la secuencia completa de estados ocultos en el codificador utilizando una técnica de agrupación. Luego, se recupera la representación completa de la secuencia en el decodificador mediante una técnica de muestreo ascendente [[48](#_ENREF_48)].

## Técnicas de ensamble de clasificadores.

La técnica de ensamble o aprendizaje conjunto, se refiere a métodos que integran múltiples modelos de referencia en el mismo marco para obtener un modelo más fuerte. El éxito de un método de ensamble depende de varios factores, incluido como se entran los modelos de referencia y como se combinan sus predicciones [[49](#_ENREF_49)].

Las técnicas de aprendizaje conjunto han logrado un rendimiento de vanguardia en diversas aplicaciones de aprendizaje automático al combinar las predicciones de dos o más modelos base. La idea fundamental detrás del aprendizaje conjunto es el reconocimiento de que los modelos base tienen limitaciones y cometen errores. Por lo tanto esta técnica tiene como objetivo mejorar el rendimiento de la clasificación aprovechando las fortalezas de múltiples modelos base. Los ensambles de clasificadores se pueden dividir en dos grupos: los que usan el mismo clasificador entrenado varias veces, llamados ensambles homogéneos y los que utilizan diferentes clasificadores, llamados heterogéneos [[14](#_ENREF_14)]. A continuación se explican los métodos de ensamble más utilizados.

### Método de votación (*Voting Ensemble*).

Los métodos de votación se utilizan generalmente en problemas de clasificación o regresión para mejorar el rendimiento predictivo. Estos métodos se dividen en tres categorías: votación máxima, votación promedio y votación promedio ponderada [[13](#_ENREF_13)].

### Votación máxima.

El primer y más popular método de votación es la votación máxima [[50](#_ENREF_50)], a menudo conocido como votación mayoritaria o votación dura. La idea de este método implica recopilar predicciones para cada etiqueta de clase y predecir la etiqueta de clase con la mayor cantidad de votos. Suponiendo que combinamos tres clasificadores, C1, C2, C3 que asignan las siguientes clasificaciones a una muestra de entrenamiento: [0, 0,1], el método de votación máxima estaría dando como resultado [0, 0,1]=0 ya que fue la de mayor voto. Este método tiene la ventaja de ser fácil de entender y es el método más sencillo de implementar [[13](#_ENREF_13), [51](#_ENREF_51)]. En la figura 1 se puede ver el flujo de votación por mayoría

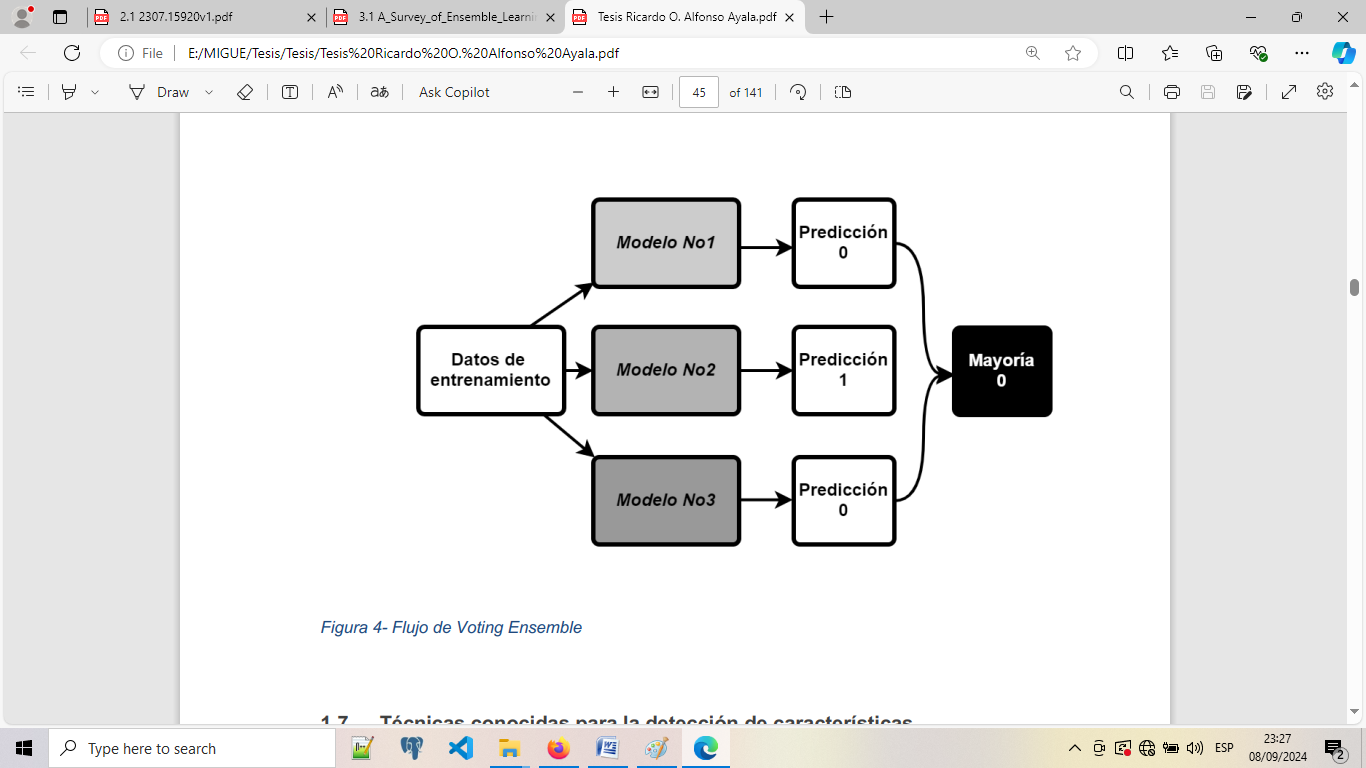


Figura 1. Flujo de ensamble por votación máxima.

### Votación promedio.

El segundo método de votación es la votación promedio [[52](#_ENREF_52)]. La idea de promediar la votación es que las predicciones se extraen de múltiples modelos y se utiliza un promedio de las predicciones para hacer la predicción final. La predicción promedio se calcula utilizando la media aritmética, que es la suma de las predicciones dividida por el total de predicciones realizadas. Por ejemplo, supongamos que el conjunto de clasificadores contiene tres miembros: C1(x)= [0.9, 0.1], C2(x)= [0.2, 0.8] y C3(x)= [0.6, 0.4]. La predicción media sería la siguiente: para calcular la clase 0, [0,9 + 0,2 + 0,6/3] = 0,566. Y calcular la clase 1, [0,1 + 0,8 + 0,4 /3] = 0,433 arrojaría una predicción y = 0. El método de votación promedio tiene la ventaja de ser el más fuerte desde el punto de vista del poder predictivo. Además, su rendimiento es más preciso que la votación por mayoría y reduce el sobreajuste. Las desventajas del método de votación promedio incluyen que es computacionalmente más costoso que el método de votación máxima, ya que requiere promediar los resultados de predicción de todos los modelos de referencia [[13](#_ENREF_13)].

### Votación promedio ponderada.

El tercer método de votación es la votación promedio ponderada, que es una versión ligeramente modificada del voto promedio [[53](#_ENREF_53)]. La idea de la votación promedio ponderada consiste en otorgar diferentes pesos a los alumnos de referencia, lo que indica la importancia de cada modelo en la predicción. Después se multiplica cada predicción por el peso de los clasificadores para producir una suma ponderada y luego dividiendo el resultado por la suma de los pesos del clasificador, estos pesos se pueden usar para calcular el promedio ponderado para cada clase 0 o clase 1. Por ejemplo, supongamos que el conjunto de clasificadores contiene tres miembros: C1(x)= [97.2, 2.8], C2(x)= [100.0, 0] y C3(x)= [95.8, 4.2]. Tiene pesos constantes para los miembros del conjunto [0,84, 0,87, 0,75]. Para calcular la clase 0 = ((97.2 \* 0.84) + (100.0 \* 0.87) + (95.8 \* 0.75))/ (0.84 + 0.87 + 0.75) = 97.763. Y para calcular la clase 1 y 1 = ((2.8 \* 0.84) + (0 \* 0.87) (4.2 \* 0.75))/ (0.84 + 0.87 + 0.75) = 2.235, se obtendría una predicción y = 0. Este método de votación es más preciso que el método de votación promedio simple. El desafío al utilizar un conjunto promedio ponderado es elegir la ponderación relativa de cada miembro. Además, el cálculo es más caro que el método de votación promedio, ya que requiere calcular el promedio ponderado de los resultados de predicción de todos los modelos de referencia, lo que lo hace de poca aplicación [[13](#_ENREF_13)].

### Método de metaaprendizaje.

El metaaprendizaje, también conocido como "aprender a aprender", es un proceso que implica aprender de los propios aprendizajes [[54](#_ENREF_54)]. Este término abarca el aprendizaje basado en experiencias previas con otras tareas. Se utiliza para mejorar el rendimiento y los resultados de un algoritmo de aprendizaje, ajustando ciertos aspectos del algoritmo según los resultados de los experimentos. El metaaprendizaje se diferencia de los modelos tradicionales de aprendizaje automático al incorporar múltiples etapas de aprendizaje. Los resultados de los inductores individuales sirven como entrada para un metaaprendiz, que finalmente genera el resultado final. En los últimos años, el interés en el metaaprendizaje ha aumentado considerablemente, especialmente después de 2017. El auge de algoritmos avanzados de aprendizaje automático ha planteado nuevos desafíos para su entrenamiento, lo que ha impulsado la investigación en metaaprendizaje. Estos algoritmos enfrentan dificultades como altos costos operativos debido a la gran cantidad de experimentos durante la fase de entrenamiento, y la dificultad de encontrar el modelo óptimo para un conjunto de datos determinado [[13](#_ENREF_13), [51](#_ENREF_51)]

### Método de embolsado (*Bagging*).

El método de bagging [[20](#_ENREF_20)], es un algoritmo completamente específico de los datos. Se basa en la creación de múltiples subconjuntos de datos a partir del conjunto de datos original. El objetivo del bagging es generar modelos predictivos más diversos ajustando la distribución estocástica de los conjuntos de entrenamiento, de manera que pequeñas variaciones en el conjunto de entrenamiento produzcan cambios significativos en las predicciones del modelo. Bagging es una abreviatura de la combinación de bootstrapping y agregación. En el bootstrapping, el entrenamiento de los modelos del conjunto se realiza sobre réplicas bootstrap del conjunto de entrenamiento. En la agregación, el resultado final se obtiene mediante una votación mayoritaria de las predicciones de los modelos para determinar la predicción fina [[13](#_ENREF_13)].

El bagging ofrece la ventaja de reducir la varianza, lo que ayuda a evitar el sobreajuste. También funciona bien con datos de alta dimensionalidad. La desventaja del bagging es que es computacionalmente costoso y tiene un alto sesgo, además de que puede dificultar la interpretabilidad del modelo. Un buen ejemplo de bagging es el algoritmo de Bosques Aleatorios [[13](#_ENREF_13)].

### Método de apilamiento (*Stacking*).

El método de *stacking*, es una técnica de ensamblaje de modelos que se utiliza para combinar información de múltiples modelos predictivos y generar un nuevo modelo (meta-modelo). La arquitectura de un modelo de stacking involucra dos o más modelos base, conocidos como modelos de nivel 0, y un meta-modelo que combina las predicciones de los modelos base, conocido como modelo de nivel 1. En los modelos de nivel 0 (modelos base), los modelos se ajustan a los datos de entrenamiento y sus predicciones se compilan. Sin embargo, en el modelo de nivel 1 (meta-modelo), el modelo aprende cómo combinar las predicciones de los modelos base de la mejor manera. Las salidas de los modelos base que se utilizan como entrada para el meta-modelo pueden ser valores de probabilidad o etiquetas de clase en el caso de la clasificación [[13](#_ENREF_13)].

El método de stacking generalmente supera el rendimiento de todos los modelos entrenados individualmente. El sobreajuste es un problema importante con el stacking de modelos, ya que hay muchos predictores que predicen el mismo objetivo y se fusionan. Además, el stacking de múltiples niveles es costoso en términos de datos (se necesita una gran cantidad de datos para el entrenamiento) y de tiempo (ya que cada capa agrega múltiples modelos) [[13](#_ENREF_13)].

## Métodos de evaluación.

Para entrenar y evaluar modelos de ABSA, es fundamental contar con datasets cuidadosamente etiquetados. Estos datasets sirven como referencia para comparar el rendimiento de diferentes modelos y permitir una evaluación precisa de su desempeño en tareas como la extracción de aspectos y la clasificación del sentimiento asociado a cada uno. Los datasets utilizados en ABSA suelen estar diseñados para abordar diferentes subtareas. Por ejemplo, para la extracción de aspectos, los datasets se etiquetan manualmente con los aspectos relevantes extraídos de las reseñas. En cambio, para la subtarea de clasificación de sentimiento por aspecto, los datasets se etiquetan con la polaridad (positiva, negativa o neutral) asociada a cada aspecto. Es importante destacar que la calidad de los datasets utilizados en ABSA juega un papel crucial en la calidad de los modelos entrenados. Un dataset bien diseñado, con anotaciones precisas y un tamaño adecuado, permitirá obtener resultados más robustos y confiables.

### Dataset.

El análisis de sentimientos basado en aspectos se apoya en conjuntos de datos anotados para su desarrollo. Entre los más utilizados [[20](#_ENREF_20), [21](#_ENREF_21), [30-32](#_ENREF_30)] se encuentran los proporcionados por SemEval 2014 y 2016. SemEval, una serie de talleres internacionales de investigación de procesamiento de lenguaje natural, promueve el avance en el análisis semántico y la creación de conjuntos de datos anotados de alta calidad. De los datasets de 2014 y 2016, los más utilizados son los relacionados con hoteles, restaurantes y laptops, disponibles en varios idiomas. Estos conjuntos de datos están etiquetados con la estructura E#A, donde E representa la entidad (restaurante, servicio, comida) y A el aspecto (precio, calidad) al que se refiere la opinión. Cada etiqueta incluye la polaridad asociada a la entidad y el aspecto [[55](#_ENREF_55)].

Los dataset encontrados en la literatura para la tarea de Análisis de Sentimientos Basado en Aspecto para el idioma español son:

* TASS 2014 SocialTV: Este conjunto de datos contiene aproximadamente 2773 tuits en español, centrados en la final de la Copa del Rey en España. A pesar de estar en español, las etiquetas abarcan las tareas de detección de la categoría del aspecto y clasificación del sentimiento del aspecto [[31](#_ENREF_31)].
* Semeval 2016 Restaurant Reviews Spanish Corpus: Está compuesto por reseñas etiquetadas con su polaridad (positiva, negativa o neutral) y con los aspectos a los que se refieren (comida, servicio, precio, ambiente, entre otros) en el contexto de la tarea 5 de SemEval-2016. Este corpus contiene aproximadamente 1000 reseñas etiquetadas manualmente [[55](#_ENREF_55)].

### Métricas utilizadas

**Las métricas son valores numéricos que miden el rendimiento de un modelo de aprendizaje automático cuando se trata de asignar observaciones a ciertas clases.** Generalmente, el rendimiento se presenta en un rango de 0 a 1, donde la puntuación 1 corresponde a un modelo perfecto, clasificado correctamente [[17](#_ENREF_17)].

En la actualidad existen un gran número de métricas, pero a continuación solo se describen las métricas utilizadas en este trabajo las cuales fueron una recopilación de las más usadas en las bibliografías.

Para entender el correcto funcionamiento de cada una de las métricas es fundamental conocer los siguientes 4 conceptos [[44](#_ENREF_44)]:

**Verdaderos Positivos (VP):** Se refieren cuando la clasificación es positiva y la predicción realizada también lo es.

**Verdaderos Negativos (VN):** Cuando la clasificación es negativa y el modelo también predice negativa.

**Falsos Positivos (FP):** Se refiere cuando el modelo realiza una predicción de positivo, cuando realmente era negativo, es decir el modelo se equivoca.

**Falsos Negativos (FN):** Cuando el modelo predice negativo cuando realmente es positivo, en este caso también se equivoca.

### Precisión

Es la razón entre el número de documentos clasificados correctamente como pertenecientes a la clase A y el número total de documentos de que han sido clasificados por el modelo como de clase A. La precisión mide la proporción de identificaciones positivas que son realmente correctas [[44](#_ENREF_44), [56](#_ENREF_56)].

**Precisión = 𝑇𝑃 / 𝑇𝑃 + 𝐹𝑃**

### Recall

La métrica Recall (Exhaustividad) es la relación entre los documentos clasificados correctamente como pertenecientes la clase A y la suma de todos los documentos de la clase A [[44](#_ENREF_44), [56](#_ENREF_56)].

**Recall = 𝑇𝑃 / 𝑇𝑃 + 𝐹𝑁**

### F1-Score

El valor F1 se utiliza para combinar las medidas de precisión y recall en un sólo valor. Esto es práctico porque hace más fácil el poder comparar el rendimiento combinado de la precisión y la exhaustividad entre varias soluciones [[44](#_ENREF_44), [56](#_ENREF_56)].

F1 se calcula haciendo la media armónica entre la precisión y la exhaustividad:

**𝐹1 = 2 ∗ (𝑝𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛 ∗ 𝑟𝑒𝑐𝑎𝑙𝑙 /𝑝𝑟𝑒𝑐𝑖𝑠𝑖𝑜𝑛 + 𝑟𝑒𝑐𝑎𝑙l)**

### Accuracy

La exactitud (Accuracy) mide el porcentaje de casos que el modelo ha acertado. Esta es una de las métricas más utilizadas. El problema con la exactitud es que nos puede llevar al engaño, es decir, puede hacer que un modelo malo parezca que es mucho mejor de lo que es. Esto se debe a que el Accuracy valora el rendimiento del modelo teniendo en cuenta todas las clases del sistema y si estas están desbalanceadas los valores no van a ser confiables [[44](#_ENREF_44), [56](#_ENREF_56)].

**Accuracy = (𝑇𝑃 + 𝑇𝑁) / (𝑇𝑃 + 𝑇𝑁 + 𝐹𝑃 + 𝐹𝑁)**

## Conclusiones parciales.

Luego del análisis de lo planteado en este capítulo se llegaron a las siguientes conclusiones:

* Ha existido en los últimos años un aumento por parte de los investigadores en el área del Análisis de Sentimientos basado en Aspectos.
* Realizar la extracción de aspectos de las opiniones en español sigue siendo un gran desafío, ya que la mayoría de las soluciones que reportan buena tasa de eficiencia se encuentra en inglés.
* Se aprecia que de los métodos utilizados para el análisis de sentimientos el más utilizado es el método supervisado ya que ha sido el que ha mostrado mejores resultados.
* Las soluciones reportadas en la literatura, que emplean aprendizaje supervisado, precisan la existencia de suficientes datos para entrenamiento, los cuales existen muy pocos para idiomas como el español.
* Se analizaron los modelos de *Deep Learning* que utilizan la estructura *Transformers*, estos moldeos pre-entrenados son los que ofrecen mejores resultados en la actualidad*.*
* Se describieron las métricas más utilizadas en la bibliografía para valorar la efectividad y la calidad de los algoritmos automáticos.
* Se observaron los aspectos generales de la investigación y su marco teórico.

# Capítulo 2: Solución de análisis de sentimientos basado en aspectos para idioma español.

Este capítulo presenta una descripción detallada de la solución propuesta para el análisis de sentimientos basado en aspectos, incluyendo los principales conceptos y entidades que la conforman, así como un esquema general que ilustra su diseño e implementación. Se expondrá, paso a paso, el funcionamiento del método propuesto, brindando una comprensión completa de su lógica y funcionamiento interno.

## Descripción de la solución.

El método propuesto tiene como objetivo extraer los aspectos de las opiniones y determinar si estos aspectos expresan un sentimiento positivo o negativo utilizando un enfoque de aprendizaje supervisado. Este método está comprendido por tres pasos principales: (1) Preprocesamiento, (2) Extracción de aspectos y (3) Predicción de polaridad.

El algoritmo presentado fue desarrollado en el lenguaje de programación Python, usando las bibliotecas de PLN. En la figura 2 se muestra el flujo general de trabajo, donde se especifican los elementos principales de cada paso.

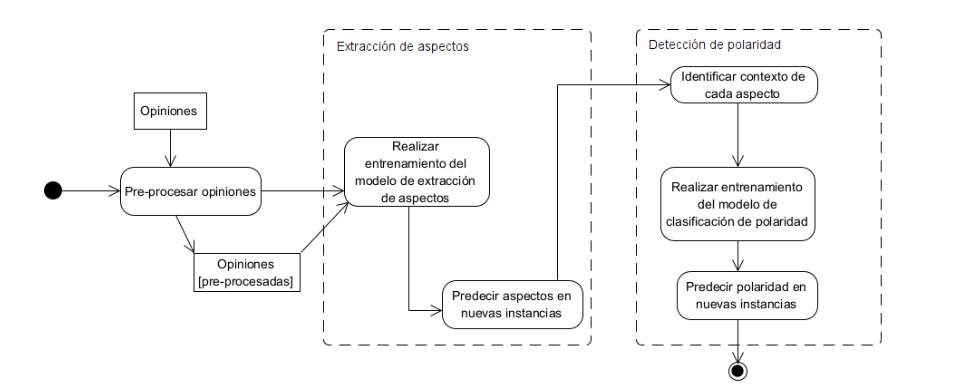


Figura 2. Esquema general del flujo del ABSA utilizando un enfoque supervisado.

### Preprocesamiento

El análisis de sentimientos basado en aspectos requiere un preprocesamiento específico para preparar los datos de texto y facilitar la extracción de información relevante en las opiniones. Si bien las tareas de preprocesamiento varían dependiendo del objetivo, en el caso del entrenamiento de los modelos *Transformers*, ciertas etapas son cruciales e indispensables, mientras que otras pueden omitirse.

Los modelos *Transformers*, han sido entrenados con millones de texto que conserva información como las mayúsculas y la puntuación. Por lo tanto, no es necesario convertir el texto a minúsculas ni eliminar los signos de puntuación, ya que el modelo ha aprendido a utilizar estas características para generar representaciones de alta calidad.

De igual manera, la eliminación de *stopwords* (palabras vacías como artículos y preposiciones) no es necesaria, dado que el corpus de entrenamiento ya lo incluye y su *tokenizador* las maneja sin problemas. En el contexto de la extracción de aspectos, una palabra como "servicio" puede ser un aspecto, pero también puede ser parte de una frase como "calidad de la comida". Eliminar las stopwords en estos casos podría implicar la pérdida de información relevante.

El primer paso en el preprocesamiento fue eliminar los caracteres especiales, dado que estos no son entendidos por los modelos, luego para la fase de entrenamiento del modelo fue necesario llevar el texto a un formato específico que fuese entendido por el modelo de lenguaje. Los datos de entrenamiento están compuestos por tres columnas, una con las opiniones, otra con el aspecto sobre el cual habla la opinión y una tercera que indica la polaridad del aspecto en dicha opinión.

En la Figura 3 se muestra como quedaría la representación para algunas reseñas. Por ejemplo, para la primera reseña: “La comida estuvo sabrosa”, el aspecto relevante es “comida” por tanto, en la columna objetivo se asigna un 1 en la posición de la palabra que constituye un aspecto a predecir y 0 en el resto de las posiciones ya que no son relevantes para la tarea. En el caso de la clasificación de polaridad la columna objetivo es un valor numérico que mapea cada una de los aspectos de la oración: no es aspecto (-1), negativo (0) o positivo (1).

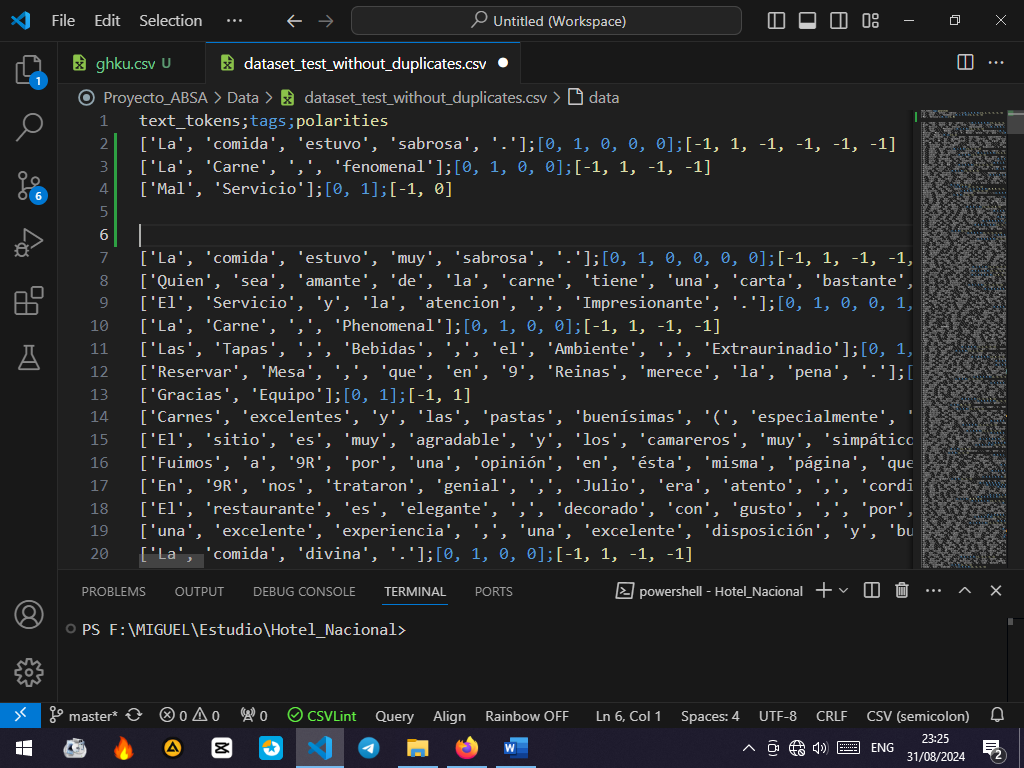


Figura 3. Representación de las reseñas.

### Extracción de aspectos.

Para abordar esta tarea, se propone un enfoque basado en un *Voting Ensemble* con votación máxima, específicamente BERT, ALBERT, RoBERTa, ELECTRA y GPT2. Este enfoque, combina los resultados de cada modelo a través de una votación de sus predicciones para lograr una mayor precisión en la identificación de aspectos.

El modelo de *Voting Ensemble* se compone de tres etapas principales: entrenamiento de modelos base, generación de predicciones y votación de predicciones. En la primera etapa, se entrenan por separado los modelos BERT, ALBERT, RoBERTa, ELECTRA y GPT2 utilizando el conjunto de datos *SemEval* 2016. Cada modelo se configura con una arquitectura específica y se optimiza para la tarea de extracción de aspectos. Una vez entrenados los modelos base, se utilizan para generar predicciones de aspectos sobre un conjunto de datos de prueba. Cada modelo produce un conjunto de predicciones, que se combinan posteriormente mediante un proceso de votación.

Para el proceso de votación, se utiliza una estrategia de mayoría simple. Se selecciona la predicción más frecuente entre los cinco modelos base para cada aspecto. Esto significa que la predicción final de aspectos se determina por la predicción que recibe la mayor cantidad de votos entre los modelos BERT, ALBERT, RoBERTa, ELECTRA y GPT2.

## Desarrollo de la solución de análisis de sentimientos basado en aspectos.

En esta sección, se detallan los elementos clave de la ingeniería de software utilizados en la construcción del sistema. Entre estos elementos se encuentra el modelo de dominio, el diagrama de casos de uso del sistema, la descripción detallada de los casos de uso, los requisitos funcionales y no funcionales y la descripción de cada uno de los paquetes y clases del software.

### Modelo de dominio.

En este capítulo se revisa el modelo de dominio, el cual es una representación visual de clases conceptuales u objetos de situaciones reales en un dominio. Los modelos de dominio también se han denominado modelos conceptuales o modelos de objetos de dominio. Los modelos de dominio son la técnica más importante en las actividades de análisis. Es fundamental entender que solo cubren conceptos que son comprensibles para el cliente y el usuario [[57](#_ENREF_57)]. Aplicando la notación UML se ilustra un modelo de dominio en la figura 4.

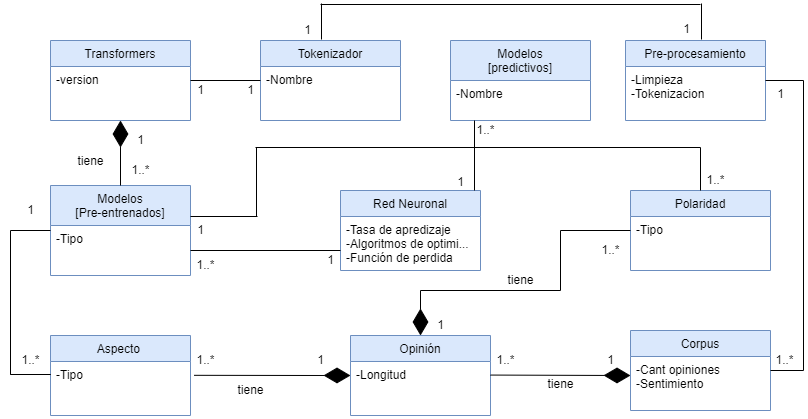


Figura 4. Modelo de dominio.

### Glosario de términos.

El siguiente glosario de términos facilita tener una comprensión más detallada de los conceptos empleados en el modelo de dominio.

Tabla 2. Glosario de término.

|  |  |
| --- | --- |
| **Nombre del concepto** | **Descripción** |
| Opinión | Texto que expresa o manifiesta el punto de vista de una persona hacia un elemento determinado |
| Corpus | Colección de textos, en este caso, de opiniones. También conocido como dataset |
| Polaridad | Representa el sentimiento hacia el aspecto, que puede ser negativo, positivo o neutro. |
| Aspecto | Representa un atributo o característica de una determinada entidad |
| Modelos pre-entrenados | Modelos entrenados con volúmenes masivos de datos, como BERT, RoBERTa, entre otros |
| Tokenizador | Función que se encarga de convertir el texto en tokens. |
| Red Neuronal | Arquitectura compuesta por distintos elementos utilizada para determinar aspectos y polaridad. |
| Preprocesamiento | Acciones de limpieza y organización de datos previo al entrenamiento y la predicción. |
| Transformers | Biblioteca que permite trabajar con distintos modelos que utilicen esta arquitectura. |
| Modelos predictivos | Entrenamiento resultante de la red neuronal. |

### Captura de requisitos.

### Requisitos funcionales.

R1: El sistema debe ser capaz entrenar un modelo de Deep Learning para extracción de aspectos.

R2: El sistema debe ser capaz de, dado un conjunto de opiniones, identificar y devolver los aspectos que los usuarios mencionan en sus comentarios.

R3: El sistema debe ser capaz de identificar la polaridad del sentimiento expresados en el texto relacionado con cada aspecto identificado.

R4: El sistema debe ser capaz de realizar limpieza del texto y correcciones gramaticales.

R5: El sistema debe permitir, dado un archivo en formato .csv o .txt, cargar los datos en memoria interna para procesarlos.

R6: El sistema debe permitir seleccionar una ruta donde se encuentra el archivo con opiniones que desea cargar el usuario.

R7: El sistema debe ser capaz de devolver el por ciento de opiniones positivas y negativas y los aspectos más mencionados dado un conjunto de opiniones.

R8: El sistema debe evaluar la calidad de la extracción de aspectos.

R9: El sistema debe evaluar la calidad de la clasificación de sentimientos.

R10: El sistema debe ser capaz de entrenar un modelo para clasificar polaridad en opiniones.

R11: El sistema debe ser capaz de crear los vectores de palabras para introducir a la red neuronal.

R12: El sistema de mostrar los resultados en forma de tablas y gráficos.

### Requisitos no funcionales.

R1: El sistema debe demostrar una alta precisión en la clasificación de aspectos.

R2: El sistema debe demostrar una alta precisión en la clasificación de la polaridad.

R3: El sistema debe ser capaz de escalar en forma horizontal.

R4: El método debe ser capaz de generalizar eficientemente a datos no vistos.

R5: El sistema debe ser capaz de adaptarse y rendir en dominios diferentes.

R6: El método debe demostrar resistencia ante ruido en los datos y ambigüedad lingüística.

R7: El sistema debe ser fácil de mantener y actualizar.

R8: El sistema debe ser fácil de usar y comprender para los usuarios.

### Diagrama de caso de uso del sistema.

El diagrama de casos de uso del sistema ofrece una visión general de las funcionalidades principales que el sistema debe proporcionar, permitiendo comprender las interacciones entre los usuarios y el sistema desde un punto de vista estático.

La Figura 5 presenta el diagrama de casos de uso desarrollado para la presente propuesta, el cual ilustra las diferentes acciones que el sistema permite realizar y las relaciones entre los distintos actores involucrados.

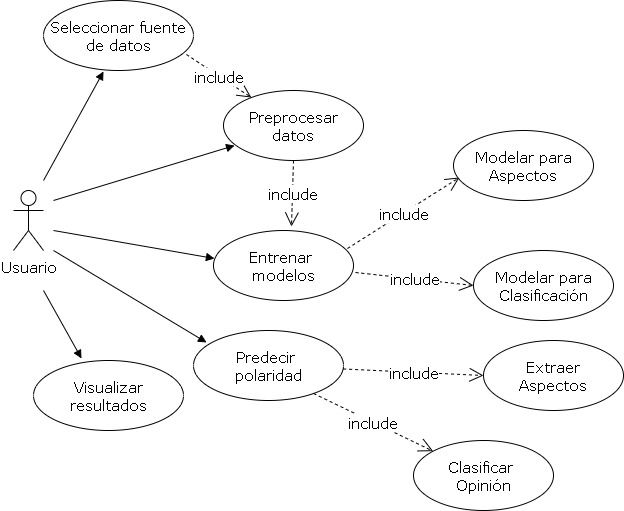


Figura 5.Diagrama de casos de uso.

### Descripción detallada de los casos de uso.

A continuación, se presenta la descripción detallada para cada caso de uso:

Tabla 3. Descripción del CUS “Seleccionar fuente de datos”.

|  |  |
| --- | --- |
| **Caso de uso** | Seleccionar fuente de datos |
| **Actores** | Usuario (inicia) |
| **Descripción** | EL caso de uso comienza cuando el usuario inicia la aplicación y busca en los archivos locales un conjunto de datos en formato .csv o .txt donde tiene las opiniones almacenadas. |
| **Requisitos funcionales** | R5, R6 |
| **Casos de uso asociados** | --- |
| **Precondición** | Debe existir un conjunto de datos local en formato csv o txt que contenga las opiniones. |
| **Postcondición** | Se cargan las opiniones para comenzar el análisis. |
| **Requisitos no funcionales** | R7, R8 |

Tabla 4. Descripción del CUS “Preprocesar datos”.

|  |  |
| --- | --- |
| **Caso de uso** | Preprocesar datos |
| **Actores** | Usuario (inicia) |
| **Descripción** | El caso de uso comienza cuando se han cargado en memoria interna las opiniones que se desean analizar, se eliminan los caracteres especiales. Además, para la fase de entrenamiento en el preprocesamiento los datos se llevan a un formato que sea comprendido por la red neuronal, es decir, se construye un vector binario, en el caso del entrenamiento para extracción de aspectos y en el caso de clasificación de polaridad se construye la ventana de palabras, se transforma la columna clase en valores numéricos, se tokenizan las reseñas, entre otros. |
| **Requisitos funcionales** | R4, R11 |
| **Casos de uso asociados** | --- |
| **Precondición** | Se debe suministrar un conjunto de datos que contenga las opiniones a analizar. |
| **Postcondición** | Se devuelven las opiniones preprocesadas. |
| **Requisitos no funcionales** | R3, R5 |

Tabla 5. Descripción del CUS “Entrenar modelos”.

|  |  |
| --- | --- |
| **Caso de uso** | Entrenar modelos |
| **Actores** | Usuario (inicia) |
| **Descripción** | El caso de uso comienza cuando se han seleccionado un conjunto de datos y este ha sido preprocesado para la tarea que se desea realizar. Luego se escoge si se va a realizar entrenamiento de un modelo de extracción de aspectos o de clasificación de polaridad y se comienza con el entrenamiento. El caso de uso termina cuando se ha obtenido el modelo entrenado para la tarea en cuestión. |
| **Requisitos funcionales** | R1, R10 |
| **Casos de uso asociados** | Modelar para aspectos <include>  Modelar para clasificación <include> |
| **Precondición** | Se debe suministrar un conjunto de datos que contenga las reseñas a analizar preprocesadas. |
| **Postcondición** | Se obtienen el modelo entrenado para la tarea seleccionada. |
| **Requisitos no funcionales** | R3, R4, R5 |

Tabla 6. Descripción del CUS “Modelar para aspectos”.

|  |  |
| --- | --- |
| **Caso de uso** | Modelar para aspectos |
| **Actores** | Usuario |
| **Descripción** | El caso de uso comienza cuando todas las reseñas han sido preprocesadas, luego comienza el proceso de entrenamiento, donde se inicializa el modelo de Deep Learning pre-entrenado combinado con la capa lineal y la función de activación, se van seleccionando lotes de datos que son pasados al modelo, se estima el error del modelo con la función de pérdida *Cross Entropy*, se propaga el error hacia atrás por la red y se vuelven a ajustar los pesos. El caso de uso termina cuando se ha entrenado el modelo todas las iteraciones que se definieron inicialmente. |
| **Requisitos funcionales** | R1, R8 |
| **Casos de uso asociados** | --- |
| **Precondición** | Se debe suministrar un conjunto de datos que contenga las reseñas preprocesadas con la columna clase con la estructura de vector binario utilizada para denotar la presencia o no de aspectos. |
| **Postcondición** | Se obtienen un modelo entrenado para la tarea de extracción de aspectos que puede ser utilizado posteriormente para realizar predicciones. |
| **Requisitos no funcionales** | R1, R3 |

Tabla 7. Descripción del CUS “Modelar para clasificación”.

|  |  |
| --- | --- |
| **Caso de uso** | Modelar para clasificación |
| **Actores** | Usuario |
| **Descripción** | El caso de uso comienza cuando todas las opiniones han sido preprocesadas, luego comienza el proceso de entrenamiento, donde se inicializa el modelo de Deep Learning pre-entrenado combinado con la capa lineal y la función de activación, se van seleccionando lotes de datos que son pasados al modelo, se estima el error del modelo con la función de pérdida *Cross Entropy*, se propaga el error hacia atrás por la red y se vuelven a ajustar los pesos. El caso de uso termina cuando se ha entrenado el modelo todas las iteraciones que se definieron inicialmente. |
| **Requisitos funcionales** | R9, R10 |
| **Casos de uso asociados** | --- |
| **Precondición** | Se debe suministrar un conjunto de datos que contenga las opiniones preprocesadas con la columna clase que contiene las etiquetas positivas o negativo. |
| **Postcondición** | Se obtienen un modelo entrenado para la tarea de clasificación de polaridad que puede ser utilizado posteriormente para realizar predicciones. |
| **Requisitos no funcionales** | R2, R3, R4 |

Tabla 8. Descripción del CUS “Predecir polaridad”.

|  |  |
| --- | --- |
| **Caso de uso** | Predecir polaridad |
| **Actores** | Usuario (inicia) |
| **Descripción** | El caso de uso comienza cuando se carga en el sistema un conjunto de datos, luego estos datos son preprocesados y una vez que están listos son pasados al módulo de extracción de aspectos donde se seleccionan los aspectos sobre los cuales se emite una opinión y se clasifica su polaridad. El caso de uso termina cuando se han extraído los aspectos de cada opinión y se ha clasificado la polaridad de cada opinión para cada uno de los aspectos. |
| **Requisitos funcionales** | R2, R3, R7 |
| **Casos de uso asociados** | Extraer aspectos <include>  Clasificar opinión <include> |
| **Precondición** | Se debe suministrar un conjunto de datos que contenga las reseñas a analizar. |
| **Postcondición** | Se obtienen por cada aspecto sus reseñas asociadas clasificadas en positivas o negativas. |
| **Requisitos no funcionales** | R1, R2, R6, R7 |

Tabla 9. Descripción del CUS “Extraer aspectos”.

|  |  |
| --- | --- |
| **Caso de uso** | Extraer aspectos |
| **Actores** | Usuario |
| **Descripción** | El caso de uso comienza cuando todas las reseñas han sido preprocesadas, luego estas pasan al proceso de extracción de aspectos para ello se carga el modelo de Deep Learning pre-entrenado y ajustado al dominio específico y se hace la predicción de los aspectos que aparecen en cada una de las opiniones. |
| **Requisitos funcionales** | R2 |
| **Casos de uso asociados** | --- |
| **Precondición** | Se debe suministrar un conjunto de datos que contenga las reseñas a analizar preprocesadas y se debe tener el modelo ajustado al dominio. |
| **Postcondición** | Se obtienen las reseñas y el o los aspectos que se encontraron en cada una de ellas. |
| **Requisitos no funcionales** | R1, R7 |

Tabla 10. Descripción del CUS “Clasificar opinión”.

|  |  |
| --- | --- |
| **Caso de uso** | Clasificar opinión |
| **Actores** | Usuario |
| **Descripción** | El caso de uso comienza cuando se han se han extraído todos los aspectos encontrados en las reseñas, luego se localizan las palabras de sentimiento asociadas a cada aspecto en los comentarios y se clasifica la polaridad para cada aspecto creando una ventana de palabras y utilizando el entrenamiento almacenado en el modelo de Deep Learning. |
| **Requisitos funcionales** | R3, R7 |
| **Casos de uso asociados** | --- |
| **Precondición** | Se debe suministrar un conjunto de datos que contenga las reseñas a analizar preprocesadas. |
| **Postcondición** | Se obtienen las reseñas por cada aspecto clasificadas en positivas o negativas. |
| **Requisitos no funcionales** | R2, R6 |

Tabla 11.Descripción del CUS “Visualizar resultados”.

|  |  |
| --- | --- |
| **Caso de uso** | Visualizar resultados |
| **Actores** | Usuario (inicia) |
| **Descripción** | El caso de uso comienza cuando se han realizados las predicciones a cada una de las opiniones y el usuario desea mostrar los resultados a través de tablas y graficas. |
| **Requisitos funcionales** | R12 |
| **Casos de uso asociados** | --- |
| **Precondición** | Se deben tener cada una de las opiniones clasificadas según su aspecto en positivas y negativas. |
| **Postcondición** | Se obtienen los resultados en forma de tablas y gráficos. |
| **Requisitos no funcionales** | R7, R8 |

### Descripción de paquetes.

A continuación, en la tabla 12 se detallan los paquetes utilizados en el proyecto.

Tabla 12. Descripción de paquetes

|  |  |
| --- | --- |
| Paquete | Descripción |
| Extracción de aspectos: | En este paquete se encuentran las funciones que permiten realizar el entrenamiento de un modelo de *Transformers* para la tarea de extracción de aspectos, así como los métodos que permiten predecir los aspectos en nuevas reseñas. |
| Data | Paquete destinado al almacenamiento del corpus sin procesar y después de ser procesado |
| Preprocesamiento: | Paquete que almacena las clases necesarias para la tarea del preprocesamiento. Algunas de esas funcionalidades fueron eliminar los caracteres especiales y llevar los datos de entrenamiento a un formato que fuese comprendido por la red. |
| Model | Paquete que almacena los modelos que fueron utilizados para entrenar, así como los modelos resultantes del entrenamiento |
| Determinar polaridad: | Este paquete contiene métodos que permiten entrenar el modelo de polaridad, así como almacenar su entrenamiento, definir sus hiperparámetros, entre otros. |
| Evaluación: | Este paquete contiene los métodos que permiten medir la calidad de la solución propuesta, así como funciones que grafican los resultados obtenidos. |

### Descripción de clases.

A continuación, en la tabla 13 se detallan las clases utilizadas en el proyecto.

Tabla 13. Descripción de las clases utilizadas.

|  |  |
| --- | --- |
| Clases | Descripción |
| Main | Clase que integra todo el flujo del método propuesto, es la encargada de darle un orden lógico a todos los pasos de ejecución y dar la salida final del algoritmo. |
| ABSAModel | Clase abstracta utilizada para definir métodos comunes tanto para el modelo de polaridad como para el de extracción de aspectos |
| AspectTermExtraction | Se utiliza para realizar, tanto entrenamiento del modelo de extracción de aspectos como predicción en nuevas instancias. |
| DatasetLoaderAspect | Clase que se encarga de cargar y preparar los datos relacionados con los aspectos para su procesamiento y entrenamiento. |
| DatasetLoaderPolarity | Similar a *DatasetLoaderAspect*, pero se enfoca en cargar y preparar datos relacionados con la polaridad de los sentimientos |
| ABSAPipeline | Representa la tubería o flujo de procesamiento que combina las diferentes etapas del análisis de sentimiento basado en aspectos, desde la identificación de aspectos hasta la clasificación de sentimientos. |
| Pytorch.nn.Module | Parte del *framework PyTorch*, es una interfaz que se utiliza para definir modelos de aprendizaje profundo. |

## Problemas frecuentes.

En el campo del análisis de sentimientos existen varios problemas frecuentes que afectan la calidad del software, la mayoría de estos afectando directamente la precisión del modelo entrenado. A continuación, se muestran algunos de los problemas más frecuentes.

### Insuficiencia de datos.

En el análisis de sentimientos una de las etapas fundamentales es la del entrenamiento del modelo. Para esto se requiere de un conjunto de datos previamente etiquetado. Si no se cuenta con suficientes datos o los que se tienen no presentan buena calidad en la clasificación, el modelo va a aprender de ellos de igual manera pero no va a presentar la precisión requerida para predecir nuevas instancias. Este problema es el problema más difícil de validar en las soluciones de análisis de sentimientos ya que se puede comprobar que el conjunto de datos supere las 2000 opiniones pero no se puede comprobar la calidad de estas. Por eso se hace fundamental seleccionar la fuente de datos para el entrenamiento de sitios oficiales.

### Desbalance de clases.

El desbalance de clase ocurre cuando se tiene una cantidad de opiniones significativas de una clase en comparación con otra. Por ejemplo si se tiene 3000 opiniones para el entrenamiento de las cuales 2000 son positivas y 1000 negativas. Esto afecta el rendimiento del modelo, ya que podría tener dificultades para capturar adecuadamente las clases minoristas, clasificando como positivas muchas opiniones que son negativas. Para evitar este problema el sistema comprueba luego de la selección de la fuente de datos que la diferencia entre una clase y la otra no supere las 300 opiniones. En caso de ser superior se muestra una alerta al usuario.

### Ruido en los datos.

El ruido en los datos es uno de los problemas que más se presentan en las opiniones en español. La mayoría de estos conjuntos de datos contienen errores ortográficos, abreviaciones, emoticones, entre otros. Estos elementos pueden afectar la precisión del modelo y llevar a resultados pocos fiables. Para dar solución a esto se utiliza una librería de Python la cual encuentra errores ortográficos en las oraciones. Se calcula estos errores en todas las opiniones y si el promedio es superior a 5 errores por oraciones se le muestra una alerta al usuario.

### Sobreajuste (*Overfitting*).

El sobreajuste es un problema frecuente en los modelos de Deep Learning, donde el modelo se ajusta demasiado a los datos de entrenamiento, perdiendo su capacidad de generalización. Esto trae como consecuencias un rendimiento deficiente en datos nuevos y no vistos. Este se evita con la agregación de algunos parámetros en el entrenamiento, ejemplo de este es añadir una o varias capas a las redes neuronales llamadas “*dropout*”, otro parámetro importante es la tasa de aprendizaje del modelo “*Lr*” el cual ayuda a evitar el sobreajuste.

## Conclusiones parciales.

Luego del análisis de este capítulo se pueden abordar las siguientes conclusiones:

* Se describió un flujo de trabajo para la tarea de análisis de sentimientos basado en aspectos:
  + Preprocesamiento.
  + Extracción de aspectos.
  + Predicción de polaridad
* En el modelo de dominio se observa la opinión como la entidad protagonista del negocio, ya que es la entrada al método y la salida es la clasificación de la misma.
* Se evidencia el papel de los modelos *Transformers* en la contribución del desarrollo de ABSA
* Se describe la estructura del proyecto y sus casos de uso brindando una mayor comprensión de la problemática.

# Capítulo 3: Evaluación de la solución propuesta.

En este capítulo se realizará la validación experimental de la investigación. Como primer aspecto, se describen las variantes de evaluación tenidas en cuenta en la investigación. Se comparan los resultados entre los modelos y con los trabajos relacionados. Para esta comparación se utilizaron las siguientes métricas: accuracy, precisión, recall y f1-score.

## Marco de Evaluación.

En este epígrafe se describe la evaluación de la solución de la extracción de aspectos. Los experimentos se ejecutaron sobre un conjunto de datos SemEval 2016 el cual contiene opiniones en español de temática de restaurantes, donde estaban presentes aspectos como servicio, comida, ambiente, atención, entre otros.

Para esta tarea de extracción de aspectos se seleccionaron diez modelos de lenguaje, los cuales fueron BETO (BERT para idioma español), BERT multilingüe (bert-multilingual-cased), ELECTRA dos versiones (versión para español), BERTIN en dos tamaños, GPT-2-spanish, y AlBERT en tres tamaños diferentes. A diferencia de BETO, que se entrena solo en el idioma español, BERT multilingual se entrena en una gran cantidad de texto en diferentes idiomas, incluyendo español, chino, árabe, hindi, portugués, francés, entre otros.

Para cada modelo se calcularon las métricas de clasificación precisión, recall y f1. Cada modelo fue entrenado durante cinco épocas (*epochs*) y el tamaño del *batch* o lote, se definió igual a ocho. La elección del tamaño de lote estuvo motivada por varios factores, entre ellos está que los tamaños de lote grandes pueden llevar a fluctuaciones en las actualizaciones de los gradientes. Además, tamaños de lotes más grandes pueden conducir a una convergencia más rápida, pero existe el riesgo de que el modelo se ajuste demasiado a los datos de entrenamiento y no generalice bien en nuevos datos [[22](#_ENREF_22)].

Despees de realizar el entrenamiento de los 10 modelos, se procedió a combinar sus resultados utilizando la técnica de Voting Ensemble por votación máxima. En esta fase se realizaron múltiples pruebas con diferentes configuraciones del ensamble. Primero se probó realizar el ensamble con los 10 modelos, lo cual aumento en un tiempo considerado la clasificación del ensamble. Se probaron otras combinaciones con 7 modelos, con 5 modelos y con 3 modelos, siendo la combinación de los modelos BETO, BERT, ALBERT-base, ALBERT-large, ALBERT-xx-large, la que presento mejores resultados.

## Resultados y discurso.

A continuación se muestran los resultados obtenidos en la extracción de aspectos para opiniones de idioma español, usando como dataset SemEval 2016 descrito anteriormente en el capítulo 1.

### Resultados de la evaluación de la extracción de aspectos.

En la Tabla 14 se muestra la configuración de los distintos hiperparámetros definidos para la ejecución de los experimentos con cada uno de los dos modelos utilizados en la experimentación. Y en la Tabla 15 se muestran los resultados de las métricas precisión, recall y f1 durante las cinco épocas de entrenamiento de cada uno de los modelos.

Tabla 14. Configuración de los hiperparámetros de los modelos.

|  |  |
| --- | --- |
| **Hiperparámetros** | **Valor** |
| Semilla (*seed*) | 50 |
| Tasa de aprendizaje | 1e-5 |
| Cantidad de épocas (*epochs*) | 5 |
| Tamaño de lote | 8 |
| Tamaño máximo de la secuencia de palabras | 512 |

Tabla 15. Evaluación de la calidad de la extracción de aspectos en el dataset de SemEval2016

|  |  |  |  |
| --- | --- | --- | --- |
| **Modelos** | **Precisión** | **F1** | **Recall** |
| BETO | 85% | 85% | 85% |
| BERT | 85% | 84% | 83% |
| ALBERT-base | 83% | 84% | 86% |
| ALBERT-large | 85% | 86% | 87% |
| ALBERT-xx-large | 83% | 86% | **90%** |
| BERTIN-base | 78% | 80% | 83% |
| BERTIN-large | 83% | 82% | 82% |
| GPT-2 | 78% | 54% | 53% |
| ELECTRA-small | 81% | 80% | 78% |
| ELECTRA-base | 83% | 82% | 81% |
| **BETO + BERT + ALBERT-base + ALBERT-large + ALBERT-xx-large** | **86.42%** | **87.42%** | 89.50% |

### Análisis de los resultados de la extracción de aspectos.

En la Tabla 15 se muestran los resultados de evaluar los distintos modelos de lenguaje y en última fila los resultados del ensamble con los modelos: BETO, BERT, ALBERT-base, ALBERT-large, ALBERT-xx-large. Los resultados indican que el modelo ALBERT-xx-large obtiene mejores resultados que el resto de los modelos estudiados en términos de recall y F1. Sin embargo, modelos como BETO, BERT, ALBERT-large obtienen mejores resultados en términos de precisión. El ensamble de los cinco modelos muestra mejores resultados que el resto en cuanto a la métrica Precisión y F1, pero presento menor nivel de recall.

En la Tabla 16 se muestra la comparación de la solución en esta subtarea frente a otras soluciones que se han reportado en la literatura que emplean el mismo conjunto de datos y se evalúan para la misma subtarea en idioma español. Todas estas soluciones reportan sus resultados solo en términos de medida F1, sin embargo, en esta investigación también se consideró conveniente utilizar las demás métricas mencionadas al inicio de la sección, con el objetivo de tener una mejor comprensión del funcionamiento del modelo en dicha tarea. Como se aprecia, la solución propuesta obtiene mejores resultados que todas las soluciones que se encuentran reportadas en la literatura para esta subtarea en idioma español, logrando incluso ser superior en un 8% aproximadamente respecto a todas las soluciones que participaron en SemEval 2016. Estos resultados corroboran la ventaja que supone el uso de los modelos de lenguaje adaptados a tareas específicas de procesamiento de lenguaje natural y el uso de técnicas de ensamble.

Tabla 16. Comparación de la solución para la extracción de aspectos frente a otras soluciones reportadas en la literatura.

|  |  |  |  |
| --- | --- | --- | --- |
| **Modelos** | **Precisión** | **F1** | **Recall** |
| GTI/C\* |  | 68.50% |  |
| GTI/U\* |  | 68.30% |  |
| IIT-T/U\* |  | 64.30% |  |
| TGB/C\* |  | 55.70% |  |
| SemEval-2016 – baseline\* |  | 51.80% |  |
| [[39](#_ENREF_39)] |  | 73,07% |  |
| [[40](#_ENREF_40)] |  | 65,40% |  |
| [[22](#_ENREF_22)] | 76,70% | 76,70% | 81,50% |
| **BETO + BERT + ALBERT-base + ALBERT-large + ALBERT-xx-large** | **86.42%** | **87.42%** | **89.50%** |

\* Soluciones presentadas en SemEval-2016 Task 5 [[28](#_ENREF_28)]

## Conclusiones parciales.

La utilización de modelos Transformers como modelo de lenguaje, combinado con una capa lineal y una función de activación Softmax, permite obtener resultados notables para idioma español en la subtarea de extracción de aspectos.

La técnica de ensamble muestra una pequeña mejora en los resultados de los modelos base independiente.

Las arquitecturas BERT y ALBERT fueron las que mejor rendimiento mostraron en la subtarea de extracción de aspectos para idioma español.

La solución propuesta logra superar al estado del arte en la subtarea de extracción de aspectos.

# Conclusiones generales.

Como resultado de esta investigación se desarrolló un método para la extracción de aspectos con el uso de modelos Transformers se aplica un enfoque de ensamble de los clasificadores para extraer los aspectos de las opiniones de entrada. Cumpliéndose de esta forma el objetivo general propuesto, ya que:

* La revisión de la literatura revela una escasez de soluciones para el análisis de sentimientos basado en aspectos en el idioma español. Una de las causas principales de esta limitación es la falta de conjuntos de datos etiquetados adecuados para entrenar modelos en este idioma.
* La revisión exhaustiva de la literatura y el análisis de las fortalezas y limitaciones de las soluciones actuales indican que la aplicación del **aprendizaje por transferencia** ofrece resultados prometedores en lenguajes con recursos lingüísticos limitados para el análisis de sentimientos basado en aspectos.
* La utilización de los modelos de lenguaje con arquitectura Transformers permite superar el estado del arte para la subtarea de extracción de aspectos para idioma español.
* La utilización de técnicas de ensamble demuestran que, para esta tarea de extracción de aspectos con el conjunto de datos de SemEval 2016 se obtienen mejores resultados que los modelos base independientes.

# Recomendaciones.

* La recopilación de un corpus más amplio de reseñas en español, abarcado distinto dominios, permitirá explorar la aplicabilidad de la solución más allá del dominio de restaurantes donde se ha validado.
* Se sugiere explorar alternativas para la representación de los aspectos, con el objetivo de mejorar el entrenamiento de los modelos de lenguaje.
* Se recomienda investigar el impacto del entrenamiento con datos etiquetados en inglés y español, en un modelo multilingüe, para evaluar si esto puede mejorar la calidad de los resultados.
* Se recomienda extender la evaluación de la solución propuesta a un contexto real, más allá del entorno controlado de la investigación.
* Las pruebas con otros tipos de técnicas de ensamble son fundamental para lograr un estudio más profundo y evaluar sus resultados para compararlos con la técnica utilizada de votación máxima.
* Ampliar los modelos base de Transformers en busca de nuevas arquitecturas

# Referencias Bibliográficas

[1] Y. Wu, Z. Jin, C. Shi, P. Liang, and T. Zhan, "Research on the Application of Deep Learning-based BERT Model in Sentiment Analysis," *arXiv preprint arXiv:2403.08217,* 2024.

[2] B. Liu, "Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers," 2012.

[3] B. Liu, *Sentiment analysis and opinion mining*. Springer Nature, 2022.

[4] Y. Montaño Reyes, "Análisis de sentimiento de mensajes de correo electrónico bajo un enfoque de transfer learning," 2022.

[5] M. Shams, N. Khoshavi, and A. Baraani-Dastjerdi, "LISA: language-independent method for aspect-based sentiment analysis," *IEEE Access,* vol. 8, pp. 31034-31044, 2020.

[6] C. H. Miranda and E. Buelvas, "AspectSA: Unsupervised system for aspect based sentiment analysis in Spanish," *Prospectiva,* vol. 17, no. 1, pp. 87-95, 2019.

[7] A. P. Pandian, "Performance evaluation and comparison using deep learning techniques in sentiment analysis," *Journal of Soft Computing Paradigm (JSCP),* vol. 3, no. 02, pp. 123-134, 2021.

[8] D. López Ramos and L. Arco García, "Aprendizaje profundo para la extracción de aspectos en opiniones textuales," *Revista Cubana de Ciencias Informáticas,* vol. 13, no. 2, pp. 105-145, 2019.

[9] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," *Multimedia tools and applications,* vol. 82, no. 3, pp. 3713-3744, 2023.

[10] J. Jia, W. Liang, and Y. Liang, "A review of hybrid and ensemble in deep learning for natural language processing," *arXiv preprint arXiv:2312.05589,* 2023.

[11] S. Arroni, Y. Galán, X. M. Guzmán Guzmán, E. R. Núñez Valdéz, and A. Gómez Gómez, "Sentiment analysis and classification of hotel opinions in twitter with the transformer architecture," *International Journal of Interactive Multimedia and Artificial Intelligence,* 2023.

[12] A. Vaswani, "Attention is all you need," *Advances in Neural Information Processing Systems,* 2017.

[13] A. Mohammed and R. Kora, "A comprehensive review on ensemble deep learning: Opportunities and challenges," *Journal of King Saud University-Computer and Information Sciences,* vol. 35, no. 2, pp. 757-774, 2023.

[14] I. D. Mienye and Y. Sun, "A survey of ensemble learning: Concepts, algorithms, applications, and prospects," *IEEE Access,* vol. 10, pp. 99129-99149, 2022.

[15] K. Chowdhary and K. Chowdhary, "Natural language processing," *Fundamentals of artificial intelligence,* pp. 603-649, 2020.

[16] E. D. Liddy, "Natural language processing," 2001.

[17] G. J. R. Fernández, "Método no supervisado para la detección de polaridad en opiniones," *Trabajo de Diploma, Universidad Tecnológica de La Habana “José Antonio Echeverría,* vol. La Habana, Cuba, 2022.

[18] R. O. A. Ayala, "Método para la detección de polaridad en opiniones: un enfoque híbrido y aplicando ensamble de clasificadores,," *Trabajo de diploma presentado en opción del título de Ingeniero Informático, Universidad Tecnológica de La Habana “José Antonio Echeverría,* 2022.

[19] Y. S. Mehanna and M. B. Mahmuddin, "A semantic conceptualization using tagged bag-of-concepts for sentiment analysis," *IEEE access,* vol. 9, pp. 118736-118756, 2021.

[20] D. López and L. Arco, "Multi-domain aspect extraction based on deep and lifelong learning," in *Iberoamerican Congress on Pattern Recognition*, 2019: Springer, pp. 556-565.

[21] A. Karimi, L. Rossi, and A. Prati, "Adversarial training for aspect-based sentiment analysis with bert," in *2020 25th international conference on pattern recognition (ICPR)*, 2021: IEEE, pp. 8797-8803.

[22] P. Montañez Castelo, " Sistema de análisis de sentimientos basado en aspectos para idioma español," 2023.

[23] M. M. Truşcǎ and F. Frasincar, "Survey on aspect detection for aspect-based sentiment analysis," *Artificial Intelligence Review,* vol. 56, no. 5, pp. 3797-3846, 2023.

[24] J. Yi, T. Nasukawa, R. Bunescu, and W. Niblack, "Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques," in *Third IEEE international conference on data mining*, 2003: IEEE, pp. 427-434.

[25] E. M. Aboelela, W. Gad, and R. Ismail, "Ontology-Based Approach for Feature Level Sentiment Analysis," *International Journal of Intelligent Computing and Information Sciences,* vol. 21, no. 3, pp. 1-12, 2021.

[26] A. García-Pablos, M. Cuadros, and G. Rigau, "W2VLDA: almost unsupervised system for aspect based sentiment analysis," *Expert Systems with Applications,* vol. 91, pp. 127-137, 2018.

[27] F. Osisanwo, J. Akinsola, O. Awodele, J. Hinmikaiye, O. Olakanmi, and J. Akinjobi, "Supervised machine learning algorithms: classification and comparison," *International Journal of Computer Trends and Technology (IJCTT),* vol. 48, no. 3, pp. 128-138, 2017.

[28] M. Pontiki *et al.*, "Semeval-2016 task 5: Aspect based sentiment analysis," in *International workshop on semantic evaluation*, 2016, pp. 19-30.

[29] L. Fröhling and A. Zubiaga, "Feature-based detection of automated language models: tackling GPT-2, GPT-3 and Grover," *PeerJ Computer Science,* vol. 7, p. e443, 2021.

[30] Y. Mao, Y. Shen, C. Yu, and L. Cai, "A joint training dual-mrc framework for aspect based sentiment analysis," in *Proceedings of the AAAI conference on artificial intelligence*, 2021, vol. 35, no. 15, pp. 13543-13551.

[31] D. Mullick, "Studying Limitations of Generative Transformer based models for Aspect Based Sentiment Analysis," 2023.

[32] Ł. Augustyniak, T. Kajdanowicz, and P. Kazienko, "Comprehensive analysis of aspect term extraction methods using various text embeddings," *Computer Speech & Language,* vol. 69, p. 101217, 2021.

[33] L. De Mattei, G. De Martino, A. Iovine, A. Miaschi, M. Polignano, and G. Rambelli, "ATE ABSITA@ EVALITA2020: Overview of the aspect term extraction and aspect-based sentiment analysis task," in *CEUR WORKSHOP PROCEEDINGS*, 2020, vol. 2765: CEUR-WS, pp. 67-74.

[34] A. Gupta and U. Sharma, "Deep Learning-Based Aspect Term Extraction for Sentiment Analysis in Hindi," *Indian Journal of Science and Technology,* vol. 17, no. 7, pp. 625-634, 2024.

[35] H. Xu, B. Liu, L. Shu, and P. S. Yu, "BERT post-training for review reading comprehension and aspect-based sentiment analysis," *arXiv preprint arXiv:1904.02232,* 2019.

[36] H. Xu, B. Liu, L. Shu, and P. S. Yu, "Dombert: Domain-oriented language model for aspect-based sentiment analysis," *arXiv preprint arXiv:2004.13816,* 2020.

[37] H. Yan, J. Dai, X. Qiu, and Z. Zhang, "A unified generative framework for aspect-based sentiment analysis," *arXiv preprint arXiv:2106.04300,* 2021.

[38] M. Patel and C. I. Ezeife, "BERT-based multi-task learning for aspect-based opinion mining," in *International conference on database and expert systems applications*, 2021: Springer, pp. 192-204.

[39] C. Henríquez, F. Briceño, and D. Salcedo, "Unsupervised model for aspect-based sentiment analysis in Spanish," *IAENG International Journal of Computer Science,* vol. 46, no. 3, pp. 430-438, 2019.

[40] B.-C. Martínez-Seis, O. Pichardo-Lagunas, S. Miranda, I.-J. Perez-Cazares, and J.-A. Rodriguez-González, "Deep learning approach for aspect-based sentiment analysis of restaurants reviews in Spanish," *Computación y Sistemas,* vol. 26, no. 2, pp. 899-908, 2022.

[41] J. Devlin, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805,* 2018.

[42] I. C. González, "con BERT: Análisis de sentimientos en tuits”."

[43] J. F. Auquilla Vicuña and J. C. Mora Alvarez, "Diseño de un sistema prototipo de diálogo persona-máquina basado en la arquitectura BERT," 2022.

[44] H. Wang, X. Hu, and H. Zhang, "Sentiment analysis of commodity reviews based on ALBERT-LSTM," in *Journal of Physics: Conference Series*, 2020, vol. 1651, no. 1: IOP Publishing, p. 012022.

[45] K. L. Tan, C. P. Lee, K. S. M. Anbananthen, and K. M. Lim, "RoBERTa-LSTM: a hybrid model for sentiment analysis with transformer and recurrent neural network," *IEEE Access,* vol. 10, pp. 21517-21525, 2022.

[46] G. R. Narayanaswamy, "Exploiting BERT and RoBERTa to improve performance for aspect based sentiment analysis," 2021.

[47] E. T. R. Schneider, J. V. A. de Souza, Y. B. Gumiel, C. Moro, and E. C. Paraiso, "A GPT-2 language model for biomedical texts in Portuguese," in *2021 IEEE 34th international symposium on computer-based medical systems (CBMS)*, 2021: IEEE, pp. 474-479.

[48] S. Alrowili and K. Vijay-Shanker, "ArabicTransformer: Efficient large Arabic language model with funnel transformer and ELECTRA objective," in *Findings of the association for computational linguistics: EMNLP 2021*, 2021, pp. 1255-1261.

[49] E.-S. Apostol, A.-G. Pisică, and C.-O. Truică, "ATESA-B {\AE} RT: A Heterogeneous Ensemble Learning Model for Aspect-Based Sentiment Analysis," *arXiv preprint arXiv:2307.15920,* 2023.

[50] J. M. Montgomery, F. M. Hollenbach, and M. D. Ward, "Improving predictions using ensemble Bayesian model averaging," *Political Analysis,* vol. 20, no. 3, pp. 271-291, 2012.

[51] H.-C. Kim, S. Pang, H.-M. Je, D. Kim, and S. Y. Bang, "Constructing support vector machine ensemble," *Pattern recognition,* vol. 36, no. 12, pp. 2757-2767, 2003.

[52] G.-R. Latif-Shabgahi, "A novel algorithm for weighted average voting used in fault tolerant computing systems," *Microprocessors and Microsystems,* vol. 28, no. 7, pp. 357-361, 2004.

[53] C. Soares, P. B. Brazdil, and P. Kuba, "A meta-learning method to select the kernel width in support vector regression," *Machine learning,* vol. 54, pp. 195-209, 2004.

[54] L. Breiman, "Bagging predictors," *Machine learning,* vol. 24, pp. 123-140, 1996.

[55] J. V. Román, E. M. Cámara, J. G. Morera, and S. M. J. Zafra, "Tass 2014-the challenge of aspect-based sentiment analysis," *Procesamiento del Lenguaje Natural,* vol. 54, pp. 61-68, 2015.

[56] G. H. Zarate Calderon, "Análisis de sentimiento en información de medios periodísticos y redes sociales mediante redes neuronales recurrentes," 2021.

[57] T. Granollers, J. Navarro, J. Suarez, and M. Rodríguez, "Perspectivas en la interacción humano-tecnología," *Jaime Muñoz-Arteaga, César A. Collazos, Toni Granollers, Huizilopoztli Luna-García,* vol. 39, 2022.