



Universidad Autónoma del Estado de México
Unidad Académica Profesional Tianguistenco

“Método de n-gramas combinados para el análisis de sentimientos en español usando SVM y Redes Neuronales”

Tesis

Para obtener el título de
Doctor en Ciencias de la Computación

Presenta

Mtro. Cesar Alexis Estrada Palacios

Tutor académico: Dr. José Luis Tapia Fabela

Tutor adjunto: Dr. Everardo Efrén Granda Gutiérrez

Tutor adjunto: Dr. José Arturo Pérez Martínez

Tutor interno: Dr. Santiago Osnaya Baltierra

Tutor externo: Dr. Rodrigo Mendoza Frías

Tianguistenco, México.

Julio 2023



Resumen

Con los avances en Inteligencia Artificial, se ha logrado un importante progreso en la capacidad de analizar de manera masiva las opiniones de los usuarios, permitiendo determinar si son positivas o negativas. Esto supone un cambio significativo, ya que ya no es necesario que una persona califique cada opinión individualmente. El uso de la inteligencia artificial y los avances continuos en métodos de análisis sitúan este campo en una posición destacada.

Sin embargo, es importante destacar que los métodos más avanzados aún no son completamente efectivos. En su mayoría, estos métodos utilizan clasificadores como SVM (Support Vector Machines) o Redes Neuronales y se basan en diferentes tamaños de n-grama, por ejemplo, bigramas.

En este contexto, se propone un enfoque que busca mejorar la clasificación mediante el aprendizaje de diferentes tamaños de n-grama. Los experimentos realizados demuestran una mejora significativa en la precisión de la clasificación en comparación con el uso de un tamaño específico de n-grama. El enfoque combina unigramas, bigramas y trigramas de palabras para obtener los mejores resultados. La inclusión de diferentes tamaños de n-grama en el método es especialmente relevante, ya que permite que el clasificador cuente con una mayor cantidad de información para el análisis de sentimientos. Al utilizar únicamente un tamaño específico de n-grama,

se presenta una limitación en términos de la información capturada. Al combinar unigramas, bigramas y trigramas, se enriquece la representación textual y se capturan patrones más complejos y sutiles presentes en las críticas de cine.

Además, como parte de este trabajo se han realizado pruebas con diferentes variantes de SVM con el objetivo de ampliar el conocimiento en el campo de la inteligencia artificial. Al explorar estas diferentes variantes, se busca identificar la configuración óptima del clasificador SVM para el método de análisis de sentimientos. De esta manera, se contribuye al avance de la investigación académica en este campo y se proporciona información valiosa que puede ayudar a futuros investigadores a evitar la repetición de experimentos y a enfocarse en las variantes más efectivas de SVM en el análisis de sentimientos en críticas de cine.



Índice General

Contenido

Agradecimientos	1
Resumen.....	1
Índice General	i
Índice de figuras.....	iv
Índice de tablas	v
Capítulo 1 Introducción.....	6
1.1 Planteamiento del problema	11
1.1.3 Pregunta de investigación.....	13
1.1.2 Preguntas de apoyo.....	13
1.2 Objetivo general.....	14
1.3 Objetivos particulares	14
1.4 Hipótesis	15
1.4.1 Variables dependientes.....	15
1.4.2 Variables independientes.....	15
1.5 Delimitaciones del estudio	15
1.6 Limitaciones del estudio	16

1.7 Justificación	16
1.8 Organización de la tesis	17
Capítulo 2 Estado del arte.....	19
2.1 Clasificación de documentos basadas en la opinión: experimentos con un corpus de críticas de cine en español.....	19
2.2 Técnicas de clasificación de opiniones aplicadas a corpus en español	20
2.3 Tema o estilo? estudiando las características más útiles para atribución de autoría	22
2.4 Nuevo método para atribución de autoría en muestras de entrenamiento balanceadas y desbalanceadas del corpus C10	23
Capítulo 3 Método propuesto	25
3.1 Proceso para la clasificación de textos.....	25
3.2 Adquisición de datos.....	28
3.3 Análisis y etiquetado de los datos.....	29
3.4 Construcción y pesado de características	30
3.5 Pesado de características	33
3.6 Selección de características	35
3.7 Entrenamiento del modelo de clasificación	36
3.7.1 Clasificador.....	36
3.8 Evaluación del método	36
Capítulo 4 Experimentos y discusión	41
4.1 Modelo de n-gramas simple	43
4.2 Combinación de n-gramas	46
4.3 Stemming	47
4.4 Número de características.....	48
4.5 Variantes del clasificador	49

4.5 Discusión	50
Capítulo 5 Conclusiones y aportaciones	53
5.1 Conclusiones	53
5.2 Aportaciones.....	54
5.3 Trabajos futuros	55
Referencias.....	56



Índice de figuras

Figura 1 Proceso de clasificación de textos basado en (Mirończuk & Protasiewicz, 2018).....	26
Figura 2 Proceso del método propuesto.	27
Figura 3 Ejemplo de una crítica de cine del corpus Muchocine.	29
Figura 4 Ejemplo de un fragmento de texto del corpus Muchocine, sin aplicar preprocesamiento.....	30
Figura 5 Ejemplo del texto después de aplicar preprocesamiento.....	31
Figura 6 Palabras, bigramas y trigramas de palabra obtenidos a partir del texto.	33
Figura 7 Matriz de confusión (Macarena, 2016)	38



Índice de tablas

Tabla 1. Resultados logrados en la investigación de (Martínez-Cámara et al., 2011).....	21
Tabla 2 Resultado del etiquetado de datos.....	29
Tabla 3 Ejemplo de pesado de términos.....	35
Tabla 4 Hiperparámetros de SVM.	42
Tabla 5 Hyperparámetros de la Red neuronal.....	43
Tabla 6 Comparación de rendimiento entre diferentes tamaños de n-gramas de palabra utilizando SVM.....	44
Tabla 7 Comparación de rendimiento entre diferentes tamaños de n-gramas de palabra utilizando Redes Neuronales.....	45
Tabla 8 Comparación de la combinación de las bolsas n-gramas.	46
Tabla 9 Aplicación de stemming en la etapa de preprocesamiento.....	48
Tabla 10 Número de características utilizadas.....	48
Tabla 11 Comparación del rendimiento de las variantes de SVM	49
Tabla 12 Comparación del método propuesto con el estado del arte.....	51



Capítulo 1

Introducción

El mundo ha experimentado cambios drásticos con la llegada del COVID-19. Miles de negocios en México y en todo el mundo han sufrido pérdidas debido a la falta de clientes y a cuestiones relacionadas con la salud. La aglomeración de personas se ha convertido en un desafío durante esta pandemia, y como ha señalado el subsecretario de salud López Gattel, "COVID-19 llegó para quedarse". Ante esta situación, muchas empresas han buscado alternativas en línea para mantenerse en el mercado.

Un ejemplo claro del éxito que se puede lograr en el mundo digital es Amazon, que ha sido reconocida como la marca más valiosa del mundo desde 2018, según el informe (*Top Brand Values Report for Global 500 2020* | *Brandirectory*, 2020). Es cierto que "el futuro de las empresas está en internet", como mencionó (Forbes Staff, 2016). Sin embargo, no basta con simplemente crear una plataforma o un sitio web para el negocio. Las redes sociales también desempeñan un papel fundamental en el éxito de estas marcas.

Según (*The Sprout Social Index, Edition VIII*, 2016), el 86% de los usuarios en redes sociales siguen a las marcas en sus perfiles y el 75% de los usuarios han realizado compras basadas en lo que han visto en estas plataformas. Tener una buena imagen de marca es crucial para alcanzar el

éxito en el mundo empresarial, ya que los comentarios de los clientes y usuarios pueden mejorar o dañar la reputación de la empresa, lo que a su vez afecta su valor (Kim et al., 2015; Luo et al., 2013; Tamrakar et al., 2018). Un ejemplo reciente es la película "Lightyear", que recibió críticas negativas en las redes sociales, lo que contribuyó a su fracaso en taquilla, según informa (Crow, 2022).

Analizar estas críticas es esencial para comprender las razones detrás del fracaso de la película. Sin embargo, debido a la gran cantidad de datos que se reciben diariamente, es necesario recurrir al uso de la inteligencia artificial para automatizar el proceso. Esta tarea específica se conoce como "análisis de sentimientos" y consiste en determinar la polaridad de un texto, es decir, el sentimiento expresado en dicho texto. Los tres tipos principales de polaridad son positivo, negativo y neutral, como se menciona en el estudio de (Can et al., 2018).

El análisis de sentimientos desempeña un papel fundamental en la mejora de la toma de decisiones en las organizaciones al recopilar y analizar opiniones. En un mundo cada vez más conectado y digital, las personas expresan sus opiniones y emociones en diversas plataformas, como redes sociales, blogs, reseñas en línea y foros de discusión. Esta avalancha de datos de opinión presenta un desafío para las organizaciones, ya que necesitan extraer información relevante y útil de esta gran cantidad de información no estructurada.

Con la creciente popularidad de las redes sociales, esta tarea ha adquirido una gran importancia. El artículo "Sentiment Analysis and Opinion Mining" (Liu, 2012) fue uno de los primeros en popularizar este tema en el ámbito científico y sigue siendo ampliamente citado. Otro artículo relevante es el de (Tripathi et al., 2015), que utilizó el software Rapidminer para analizar sentimientos, lo que facilita la experimentación con diversas

herramientas de inteligencia artificial y estadística, y permite replicar el método propuesto sin la necesidad de programar, ya que ofrece una amplia gama de herramientas, métodos y modelos como componentes, lo que facilita la implementación e investigación.

En la actualidad, existen dos enfoques principales para el análisis de sentimientos: el uso de lexicones y el uso de técnicas de aprendizaje automático (machine learning). Los lexicones son diccionarios de palabras que contienen información sobre la polaridad de cada una, lo que permite determinar la polaridad de un texto completo (Abdulla et al., 2013). Por otro lado, las técnicas de aprendizaje automático requieren un conjunto de datos de entrenamiento para evaluar el sentimiento de un texto (Can et al., 2018) (Ge et al., 2017). Entre las técnicas de aprendizaje automático más efectivas, las Redes Neuronales se consideran las más poderosas (Abambres et al., 2018). Estas redes matemáticas se basan en el funcionamiento del cerebro y consisten en una red interconectada de neuronas. Reciben un conjunto de datos de entrada con su respectiva salida para generar un modelo matemático capaz de proporcionar respuestas a futuras entradas desconocidas. Aunque las Redes Neuronales no son una novedad, en el pasado requerían hardware costoso y avanzado debido a la gran cantidad de neuronas y la información que debía procesarse en cada una de ellas. Sin embargo, gracias a los avances tecnológicos, como el uso de GPU, que cuenta con miles de núcleos de procesamiento en comparación con la CPU, y el uso del cómputo en la nube (Google, 2020), este problema se ha reducido significativamente, lo que ha impulsado el uso del Deep Learning para el aprovechamiento de las Redes Neuronales en general. En resumen, el Deep Learning utiliza millones o billones de neuronas en la red neuronal para un aprendizaje más profundo.

La investigación en torno al análisis de sentimientos ha tenido un enfoque predominantemente en el idioma inglés, dejando a otros idiomas como el español con escasez de estudios en esta área (Miranda & Guzmán, 2017). Este es un problema importante, ya que gran parte de la población mundial y prácticamente todo un continente habla este idioma. Por lo tanto, es crucial contar con métodos eficaces para el análisis de sentimientos en español (de Albornoz Cuadrado, 2011). Con este trabajo, se busca desarrollar un método que logre superar el rendimiento actual en la tarea de análisis de sentimientos en críticas de cine en el idioma español.

La metodología utilizada en este trabajo sigue el proceso común en la clasificación de textos, consta de las etapas de adquisición de datos, análisis y etiquetado, construcción y pesado de características, selección de características, entrenamiento del modelo de clasificación y evaluación. Sin embargo, se presta especial atención a las etapas de construcción de características y al modelo de clasificación, con el objetivo de obtener resultados óptimos en la tarea de análisis de sentimientos en críticas de cine en español.

En la etapa de construcción de características, se propone utilizar una técnica de representación de texto conocida como "bolsa de palabras". Sin embargo, a diferencia de la bolsa de palabras tradicional en la que se utiliza solo un tamaño de n-grama, en este enfoque se combina distintos tamaños de n-gramas para formar la bolsa de palabras. El objetivo de esta combinación es permitir que el clasificador aprenda de diferentes tipos de n-gramas, en lugar de estar limitado a uno específico.

Esta técnica de combinar n-gramas ha sido aplicada previamente en tareas de atribución de autoría, como se evidencia en el trabajo de (Sari et al., 2018). En dicho estudio, se utilizaron los 100 n-gramas de palabra más frecuentes de tamaños 1, 2 y 3, lo que generó un total de 300 n-gramas.

Este enfoque demostró mejorar significativamente el proceso de clasificación en la tarea de atribución de autoría.

Además, otro estudio de atribución de autoría realizado por (Estrada & Fabela, 2022) también corroboró los beneficios de combinar n-gramas para mejorar los resultados cuando se utiliza un clasificador SVM. A pesar de que existe un trabajo previo que ha obtenido los mejores resultados en críticas de cine en español utilizando SVM, hasta el momento no se ha evaluado el impacto de la combinación de n-gramas en el análisis de sentimientos. Por esta razón, se plantea la necesidad de investigar si esta técnica de representación puede superar los resultados previos y mejorar el análisis de sentimientos en críticas de cine en español. En resumen, se propone combinar los n-gramas de tamaño 1, 2 y 3 para formar una sola bolsa de palabras y evaluar su efectividad.

Para el entrenamiento del modelo de clasificación, se propone utilizar SVM y Redes Neuronales, debido a su gran efectividad y popularidad como algoritmos de machine learning (Khanna et al., 2015). Además, se plantea evaluar distintas variantes de SVM, ya que en el trabajo de (Estrada & Fabela, 2022) se mostró que SVM *Primal* tenía un rendimiento superior al SVM estándar, lo que ayudaría a mejorar los resultados obtenidos.

Una vez completada la fase de experimentación, los resultados indican que la combinación de diferentes tamaños de n-gramas en una sola bolsa de palabras mejora la precisión de la clasificación al utilizar SVM. Específicamente, se observó que la combinación de unigramas, bigramas y trigramas de palabra obtuvo el mejor rendimiento. Aunque las distintas variantes de SVM tuvieron un impacto moderado en la clasificación, esta pequeña mejora resulta significativa en comparación con los resultados del estado del arte existente.

1.1 Planteamiento del problema

El análisis de sentimientos relacionado con las críticas de cine desempeña un papel fundamental al permitir comprender la opinión y la percepción del público en relación con las películas. La identificación y clasificación de los sentimientos expresados en estas críticas nos brinda la capacidad de evaluar la satisfacción, la insatisfacción, el entusiasmo, la decepción y otras emociones asociadas a la experiencia cinematográfica. La información recopilada resulta sumamente valiosa para la industria del cine, ya que ayuda en la toma de decisiones estratégicas, como la planificación de estrategias de marketing, la mejora de la calidad de las películas y la identificación de aspectos destacables o que requieren mejoras.

La importancia social de esta investigación radica en su capacidad para brindar una comprensión profunda de las preferencias y percepciones del público en relación con las películas en el idioma español. Esto no solo beneficia a la industria del cine, sino también a la sociedad en su conjunto. Al contar con métodos más precisos y contextualmente adecuados para el análisis de sentimientos en español, se pueden tomar decisiones más informadas y acertadas en diversos ámbitos.

Además, el avance de la tecnología y la globalización ha transformado la forma en que buscamos y compartimos opiniones sobre películas. Antes solíamos depender principalmente de las opiniones de personas cercanas, amigos o familiares, para tomar decisiones sobre qué películas ver. Sin embargo, en la era de Internet y las redes sociales, hemos ampliado nuestro alcance y ahora tenemos acceso a una gran cantidad de opiniones y reseñas de personas de diferentes partes del mundo.

Este cambio hacia una opinión más globalizada ha sido facilitado por plataformas en línea que permiten a los usuarios compartir sus experiencias

y opiniones sobre películas. Ahora podemos explorar diferentes sitios web, foros, redes sociales y aplicaciones dedicadas a las críticas de cine, lo que nos brinda una perspectiva más amplia y diversa. Este acceso a una variedad de opiniones nos ayuda a tomar decisiones informadas sobre qué películas ver y qué esperar de ellas.

En este contexto, la inteligencia artificial y el aprendizaje automático desempeñan un papel fundamental. Estas tecnologías permiten procesar grandes volúmenes de datos generados por los usuarios en forma de reseñas, comentarios y calificaciones. Al aplicar técnicas de análisis de sentimientos y aprendizaje automático, la inteligencia artificial puede ayudarnos a comprender y extraer patrones de opinión de esta enorme cantidad de información disponible en línea. Esto nos permite obtener una visión más completa y representativa de las preferencias y percepciones del público en general.

Aunque existen numerosos métodos en el estado del arte para el análisis de sentimientos, es importante destacar que la mayoría de ellos están diseñados específicamente para el idioma inglés. Estos métodos han sido ampliamente estudiados y aplicados en el contexto de la lengua inglesa, aprovechando las características particulares de dicho idioma. Sin embargo, el idioma español, siendo uno de los idiomas más hablados en el mundo, con una amplia comunidad de hablantes que supera los 493 millones de personas (Cervantes, 2021), presenta particularidades lingüísticas y culturales que requieren un enfoque específico en el análisis de sentimientos en el contexto de las críticas de cine.

Es de vital importancia desarrollar métodos específicos, precisos y contextualmente adecuados para el análisis de sentimientos en español. La aplicación directa de los métodos diseñados para el inglés resulta insuficiente y poco efectiva debido a las diferencias en vocabulario,

gramática y expresiones idiomáticas entre ambos idiomas. Por lo tanto, es necesario llevar a cabo investigaciones y desarrollar enfoques que consideren las sutilezas y matices propios de la expresión de sentimientos en español, con el objetivo de obtener resultados más precisos y relevantes en el análisis de las opiniones del público hispanohablante sobre las películas. Además, esta mejora en la precisión y relevancia de los resultados contribuirá a proporcionar información valiosa para aplicaciones en áreas como la inteligencia artificial, la toma de decisiones empresariales y el análisis de opiniones en redes sociales, entre otros.

En el contexto del análisis de sentimientos en las críticas de cine en español, se han publicado dos artículos científicos en los últimos años (Cruz et al., 2008; Martínez-Cámara et al., 2011), y el corpus Muchocine se destaca como la única referencia disponible en español. El método con mejores resultados hasta el momento logra una precisión del 87.73% mediante el uso de SVM (Martínez-Cámara et al., 2011). Por lo tanto, la pregunta de investigación planteada es:

1.1.3 Pregunta de investigación

¿Cómo desarrollar un método de análisis de sentimientos para críticas de cine en español que logre una precisión mayor al 87.73% y supere los métodos del estado del arte?

1.1.2 Preguntas de apoyo

1. ¿Cómo afecta el uso de diferentes combinaciones de n-gramas en el rendimiento del análisis de sentimientos en críticas de cine para el idioma español utilizando SVM?
2. ¿Existe una mejora significativa en el rendimiento del análisis de sentimientos en críticas de cine en español al combinar n-gramas en comparación con no combinar?

3. ¿Cuál de los clasificadores, SVM o Redes Neuronales, ofrece mejores resultados en el análisis de sentimientos en críticas de cine para el idioma español?

1.2 Objetivo general

Desarrollar un método de análisis de sentimientos para críticas de cine en español que supere la máxima precisión alcanzada en el estado del arte, la cual es del 87.73% con la finalidad mejorar la clasificación de críticas de cine del corpus Muchocine y superar los enfoques actuales.

1.3 Objetivos particulares

- 1- Identificar los desafíos y limitaciones actuales en el análisis de sentimientos en críticas de cine en español.
- 2- Revisar y analizar la literatura existente en el campo de análisis de sentimientos y procesamiento del lenguaje natural para construir un marco teórico que guíe la investigación.
- 3- Realizar un análisis detallado de los métodos y técnicas actuales utilizados en el análisis de sentimientos, con énfasis en aquellos relacionados con el corpus Muchocine.
- 4- Experimentar con diversas combinaciones de técnicas de extracción, representación y selección de características para mejorar la precisión en el análisis de sentimientos.

- 5- Optimizar los hiperparámetros de los clasificadores SVM y Redes Neuronales mediante técnicas de búsqueda empírica.
- 6- Evaluar y comparar los resultados obtenidos con los métodos actuales del estado del arte y formular conclusiones y recomendaciones para futuras investigaciones.
- 7- Diseñar un protocolo para evaluar el desempeño del método desarrollado mediante experimentos controlados con el corpus Muchocine.

1.4 Hipótesis

El uso de técnicas avanzadas de extracción de características, tales como el uso de combinaciones de n-gramas, junto con el uso de clasificadores avanzados como SVM y Redes Neuronales, permitirá superar el umbral del 87.73% de precisión, que actualmente se considera el resultado más alto del estado del arte para los métodos utilizados en el corpus Muchocine, en el contexto del análisis de sentimientos en críticas de cine en español.

1.4.1 Variables dependientes

Métricas: *Precision, Recall, Accuracy, F-Measure y Kappa.*

1.4.2 Variables independientes

N-gramas combinados, tamaño de la muestra, parámetros de SVM y la red neuronal.

1.5 Delimitaciones del estudio

- 1- El enfoque del análisis se centra exclusivamente en la polaridad de las críticas, es decir, si son positivas, negativas.
- 2- Se emplean técnicas de procesamiento del lenguaje natural y aprendizaje automático, específicamente se experimenta con distintos

métodos de extracción, representación y selección de características, así como con diferentes hiperparámetros para los clasificadores SVM y Redes Neuronales.

- 3- Se realiza una comparación de los resultados obtenidos con los métodos actuales del estado del arte en el corpus Muchocine.

1.6 Limitaciones del estudio

- 1- La métrica para superar es precisión a pesar de no ser la única utilizada en la etapa de experimentación, esto debido a que es la única que se encuentra presente en todos los trabajos que hemos utilizado como referencia para comparar el método propuesto. Por otro lado, las métricas de *Recall* y *F-measure* tienen un carácter principalmente informativo y complementario en el análisis.
- 2- El corpus utilizado en este estudio es el corpus Muchocine, el cual se ha elegido debido a ser el único corpus de críticas en español disponible en la actualidad.

1.7 Justificación

El análisis de sentimientos en críticas de cine en español se enfrenta a un desafío relevante en el campo del procesamiento del lenguaje natural y el aprendizaje automático. Este estudio, se centra en el corpus Muchocine como fuente de datos para llevar a cabo el análisis. La elección de un corpus específico en español resulta fundamental, ya que permite evaluar de forma precisa el rendimiento de los métodos desarrollados en un contexto real y en el idioma objetivo.

El objetivo principal es desarrollar un método innovador que utilice técnicas avanzadas de procesamiento del lenguaje natural y aprendizaje automático para mejorar la precisión en la detección de la polaridad en las críticas de cine, es decir, si son positivas o negativas, dado que esta

información resulta de gran relevancia para comprender la opinión y percepción del público sobre las películas.

Con el propósito de alcanzar dicho objetivo, se llevarán a cabo experimentos con métodos de extracción, representación y selección de características, así como con diferentes hiperparámetros para los clasificadores SVM y Redes Neuronales. La meta es superar los métodos actuales del estado del arte en términos de precisión en el análisis de sentimientos en críticas de cine en español, tomando el corpus Muchocine como referencia.

1.8 Organización de la tesis

A continuación, se presenta la estructura de esta tesis.

- 1- En el primer capítulo, se presenta la importancia y el propósito de la investigación, se incluye una breve descripción del problema investigado, así como las preguntas de investigación y las hipótesis.
- 2- En el segundo capítulo, se especifica los métodos y técnicas relacionadas al análisis de sentimientos y clasificación de textos que se mencionan en el estado del arte y cómo están relacionadas con la investigación, además de citar estudios específicos que sirven como comparación para la investigación.
- 3- En el tercer capítulo, se incluye una descripción detallada de las herramientas y tecnologías utilizadas en el método propuesto, junto con una discusión sobre las ventajas y desventajas de cada una.
- 4- En el cuarto capítulo, se lleva a cabo la experimentación y se comparan los resultados con el estado del arte, además se presenta una discusión detallada sobre los resultados obtenidos, incluyendo los factores que podrían haber influido en los resultados y las posibles mejoras para el futuro.

- 5- En el quinto capítulo, se presentan las conclusiones generales y se incluyen recomendaciones para futuras investigaciones en el tema, que se basan en los hallazgos y limitaciones de esta investigación.



Capítulo 2

Estado del arte

En este capítulo, se presentan los trabajos más relevantes para esta investigación. A partir de su análisis, se sintetiza el método propuesto. A pesar de que el análisis de sentimientos en críticas de cine en español es un área con poco estudio, se han encontrado dos trabajos que han aplicado métodos en el idioma español utilizando el corpus Muchocine como referencia.

Sin embargo, dado que la investigación se enfoca en el análisis de sentimientos en críticas de cine, se ha buscado información en otras áreas de clasificación de texto como la Atribución de Autoría. Estos trabajos han brindado un conocimiento valioso para nuestra investigación, ya que han explorado la combinación de n-gramas en el análisis de textos.

2.1 Clasificación de documentos basadas en la opinión: experimentos con un corpus de críticas de cine en español

El trabajo de (Cruz et al., 2008) fue pionero en el campo de análisis de sentimientos en críticas de cine en español. A falta de un corpus adecuado para este propósito, los autores proponen el corpus Muchocine, compuesto por 3878 críticas de cine extraídas del sitio web con el mismo nombre. Sin embargo, el corpus no fue utilizado en su totalidad en su estudio, el autor utilizó un total de 400 críticas, compuestas por 200 críticas

positivas y 200 críticas negativas, excluyendo las críticas neutras. Esta selección se debió a la falta de recursos disponibles para procesar todas las críticas, como se menciona en el estudio realizado por el propio autor (Cruz et al., 2008). Para la optimización de parámetros, el autor dividió las 400 críticas en un conjunto de entrenamiento, que representó el 80% del total, y un conjunto de prueba, que representó el 20% restante. El mejor resultado obtenido en este trabajo fue de 77%, esto nos sirve como una base para poder comparar los resultados del método propuesto. Un dato interesante, es que el autor de este artículo utilizó bigramas de palabra, pero no se realizaron combinaciones de n-gramas como se propone en esta tesis.

2.2 Técnicas de clasificación de opiniones aplicadas a corpus en español

Este trabajo ha continuado la investigación en el corpus Muchocine y ha logrado obtener una precisión del 87.73% utilizando SVM como clasificador, lo cual representa un avance significativo en el campo. A continuación, se presentan los resultados obtenidos en esta investigación.

Tabla 1. Resultados logrados en la investigación de (Martínez-Cámara et al., 2011)

	Stop	Stem	Re_Cr			Cr			Re		
			Prec.	Recall	F1	Prec.	Recall	F1	Prec.	Recall	F1
TF-IDF	√	√	86,84%	86,67%	86,75%	85,67%	85,49%	85,58%	76,23%	76,07%	76,15%
	√		87,66%	87,60%	87,63%	86,40%	86,34%	86,37%	75,75%	75,51%	75,63%
		√	86,80%	86,64%	86,72%	85,77%	85,56%	85,66%	75,95%	75,80%	75,87%
			87,73%	87,69%	87,71%	86,18%	86,10%	86,14%	75,57%	75,37%	75,47%
TF	√	√	79,81%	79,44%	79,62%	78,47%	78,20%	78,33%	74,74%	74,51%	74,62%
	√		77,48%	77,08%	77,28%	76,12%	75,65%	75,88%	72,12%	71,92%	72,02%
		√	79,74%	79,42%	79,58%	78,16%	77,91%	78,03%	74,74%	74,56%	74,65%
			77,06%	76,65%	76,85%	75,83%	75,34%	75,58%	72,29%	72,07%	72,18%
TO	√	√	77,66%	76,89%	77,27%	74,66%	73,13%	73,89%	73,82%	73,51%	73,66%
	√		74,05%	73,00%	73,52%	72,59%	70,96%	71,77%	71,84%	71,49%	71,66%
		√	77,86%	77,09%	77,47%	74,64%	73,03%	73,83%	74,08%	73,81%	73,94%
			74,25%	73,12%	73,68%	72,45%	70,77%	71,60%	71,87%	71,49%	71,68%
BTO	√	√	84,23%	84,20%	84,21%	83,61%	83,63%	83,62%	75,24%	75,07%	75,15%
	√		83,94%	83,91%	83,92%	83,69%	83,62%	83,65%	73,53%	73,24%	73,38%
		√	84,13%	84,12%	84,12%	84,10%	84,09%	84,09%	74,63%	74,45%	74,54%
			84,11%	84,16%	84,13%	83,84%	83,78%	83,81%	73,96%	73,70%	73,83%

Como se puede observar, los mejores resultados se obtienen cuando se utiliza SVM, *TF-IDF* (del inglés *Term frequency – Inverse document frequency*), no se aplica *stemming* (consiste en extraer la raíz de una palabra) y no se eliminan las *stopwords* (palabras sin significado como: artículos, pronombres, preposiciones). Estos datos son considerados en el desarrollo del método propuesto.

Es importante mencionar que se ha incrementado el número de críticas utilizadas en comparación con trabajos anteriores (como Cruz et al., 2008), utilizando todas las críticas disponibles con polaridad positiva y negativa. En este corpus se utilizaron 1351 críticas positivas y 1274 críticas negativas, no se evaluaron las críticas neutrales.

Otro aporte importante de este trabajo es la evidencia de que los resultados se mejoran al combinar la crítica de la película con el resumen de la misma, el cual se encuentra en el mismo corpus. Esto sugiere que, en ciertos casos, la combinación de diferentes elementos puede tener un impacto positivo en los resultados.

2.3 Tema o estilo? estudiando las características más útiles para atribución de autoría

Este artículo es de la tarea de atribución de autoría, presentado por (Sari et al., 2018), experimenta con diferentes tipos de características, incluyendo la combinación de n-gramas, para determinar cuál tiene mejores resultados en la tarea de atribución de autoría en inglés. Los títulos de características se dividen en tres categorías: estilo, contenido e híbridos.

-Características de estilo: incluyen la frecuencia de letras, signos de puntuación, números y la longitud de las palabras, entre otros.

-Características de contenido: unigramas, bigramas y trigramas de las palabras más frecuentes en el texto.

-Características híbridas: combinan bigramas y trigramas de caracteres para obtener una mejor comprensión del contenido del texto.

Hay tres aspectos clave que se destacan en este trabajo. En primer lugar, se investiga la tarea de atribución de autoría, que se enfoca en identificar al autor de un texto cuya autoría es desconocida (Plakias & Stamatatos, 2008). Esta tarea forma parte de las múltiples que se encuadran dentro de la clasificación de textos, la cual también incluye el análisis de sentimientos. Por esta razón, se considera importante esta investigación, ya que se busca establecer si los hallazgos obtenidos en la

tarea de atribución de autoría son aplicables también en la tarea de análisis de sentimientos.

Como segundo punto, se destaca que en la tarea de atribución de autoría, las características de contenido tienen un mejor rendimiento que las híbridas. Los n-gramas de palabra producen resultados superiores a los n-gramas de carácter, ya que al buscar el estilo del autor no es necesario considerar el sentimiento de las palabras.

Como tercer punto destacado, en este trabajo se menciona el uso de n-gramas en la tarea de atribución de autoría. Aunque solo se realiza un pequeño experimento combinando los 100 n-gramas de palabra más frecuentes de tamaño 1, 2 y 3, se logran resultados interesantes alcanzando un 76.2% de precisión con solo 300 características, comparado con el 80.2% obtenido por (Plakias & Stamatatos, 2008) utilizando 2500 características. Esto sugiere que la combinación de n-gramas tiene un gran potencial en la tarea de atribución de autoría.

2.4 Nuevo método para atribución de autoría en muestras de entrenamiento balanceadas y desbalanceadas del corpus C10

En este trabajo de atribución de autoría para el idioma inglés, se propone utilizar una combinación de n-gramas de palabra y algoritmo SVM para mejorar la precisión en la tarea de identificar al autor de un texto. En lugar de utilizar solo 300 características como lo hizo (Sari et al., 2018), se utilizaron 2500 términos, que es el mismo número de términos utilizado por (Plakias & Stamatatos, 2008) quien obtuvo los mejores resultados hasta 2008. Los resultados de la investigación sugieren que combinar n-gramas es beneficioso para la clasificación y logra superar los resultados del estado del arte. Si en la tarea de atribución de autoría existe una mejora al combinar n-gramas, esto nos hace pensar que podría existir también una

mejora si se utiliza la combinación de n-gramas, para el análisis de sentimientos. Sin embargo, es importante mencionar que el corpus utilizado en ese estudio está en el idioma inglés, por lo que se existen dos problemas principales, el idioma y la tarea. El uso de métodos de análisis de sentimientos en idiomas diferentes al original plantea incertidumbre acerca de su efectividad. No obstante, la aplicación de métodos de análisis de sentimientos en diversos idiomas, ha arrojado resultados prometedores (Balahur & Turchi, 2012; Brooke et al., 2009; Mohammad, 2017). A pesar de que la aplicación de estos métodos en otros idiomas no ha mostrado una disminución significativa en comparación con el idioma para el cual fueron originalmente desarrollados, es importante destacar que la principal dificultad radica en la traducción de los datos de entrenamiento. Durante el proceso de traducción, se puede perder calidad y precisión en los datos, lo que perjudica la efectividad de los métodos de análisis de sentimientos.

En el caso particular del presente estudio se cuenta con un corpus originalmente en español, lo que plantea una situación favorable. Dado que no se requiere una traducción, se prevé que los resultados obtenidos por la combinación de n-gramas sean confiables y precisos en comparación con la aplicación de estos métodos en otros idiomas.

Otro hallazgo importante es que el uso de SVM *Primal* mejora los resultados en comparación con el uso del SVM estándar. Por otro lado, se descubrió que utilizar *stemming* empeora los resultados obtenidos, lo que también se observó en el trabajo de (Martínez-Cámara et al., 2011) con el cual se compararon los resultados. En este estudio se reafirma la importancia de considerar si utilizar *stemming* mejora o no los resultados.



Capítulo 3

Método propuesto

En este capítulo se presenta el detalle del método propuesto para realizar el análisis de sentimientos en críticas de cine en español. Se siguen los procesos estándar de clasificación de textos y se muestra de forma clara y concisa en un diagrama para una fácil comprensión. Además, se describen las técnicas específicas que se utilizaron en cada etapa del proceso propuesto.

3.1 Proceso para la clasificación de textos

La tarea de análisis de sentimientos es una de las aplicaciones de la clasificación de textos (Ge et al., 2017; Vijayan et al., 2017). La mayoría de los métodos de clasificación de textos siguen una serie de etapas, según lo determinado en el estudio de (Mirończuk & Protasiewicz, 2018). A continuación, se presenta este proceso.



Figura 1 Proceso de clasificación de textos basado en (Mirończuk & Protasiewicz, 2018).

Este proceso sirve como base para nuestro método; pero se incorpora una etapa adicional que emplea técnicas avanzadas de procesamiento de lenguaje natural, como la combinación de n-gramas y la evaluación de variantes de SVM y Redes Neuronales. Estas técnicas permiten capturar de manera precisa y completa las características y patrones relevantes presentes en las críticas de cine, lo que conduce a una mayor precisión y rendimiento en el análisis de sentimientos. Por lo tanto, este enfoque mejora

y optimiza el proceso de clasificación de textos al integrar técnicas avanzadas y personalizadas en el método específico propuesto, a continuación, se presenta el método propuesto y se describe cada una de sus etapas.

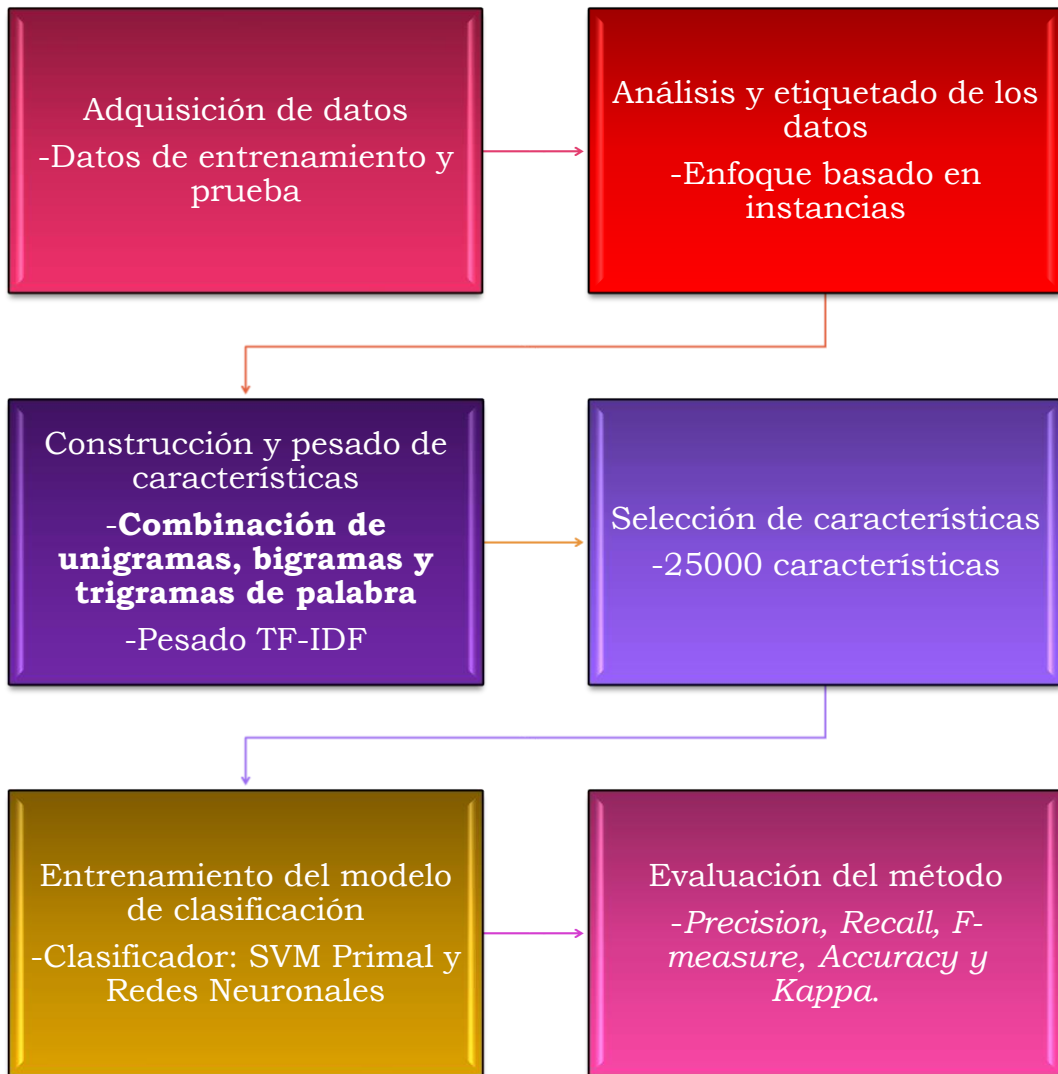


Figura 2 Proceso del método propuesto.

3.2 Adquisición de datos

Los datos son el elemento clave en el aprendizaje supervisado. Para utilizar estos métodos, es necesario tener un conjunto de entrenamiento y otro de evaluación, que pueden obtenerse de diferentes fuentes (Mirończuk & Protasiewicz, 2018). En este caso, los datos provienen del corpus Muchocine, compuesto por 3878 críticas de cine en español obtenidas del sitio web Muchocine (Cruz et al., 2008). Este corpus presenta un desafío, ya que las críticas fueron escritas por personas comunes, es decir, individuos sin experiencia o formación especializada en el ámbito de la crítica cinematográfica. En contraste, las críticas profesionales son redactadas por personas con conocimientos y experiencia específicos en el campo. Esta distinción entre ambos tipos de personas puede dificultar la clasificación de las críticas debido a la presencia de errores gramaticales, ortográficos, incoherencias, uso de lenguaje coloquial y expresiones poco comunes en las críticas escritas por personas comunes. Cada crítica está calificada con un número del 1 al 5, donde 1 significa muy mala, 2 mala, 3 neutra, 4 buena y 5 muy buena. Para fines comparativos, no se consideran las críticas con puntuación 3, tal como se ha establecido en los trabajos de referencia (Cruz et al., 2008; Martínez-Cámara et al., 2011). Las críticas con puntuaciones 1 o 2 se clasifican como negativas y las críticas con puntuaciones 4 o 5 como positivas, de acuerdo con (Cruz et al., 2008; Martínez-Cámara et al., 2011). En total, el corpus cuenta con 1274 críticas negativas y 1351 positivas, por lo que está casi balanceado.

Los datos originales de las críticas cinematográficas vienen en formato HTML, para facilitar el proceso y reducir el tamaño del corpus, se convierten a ficheros XML a través del software Rapidminer, el cual tiene un componente específico para esta tarea. Cada crítica consta de cinco elementos: autor, título de la película, resumen, texto completo de la crítica

y la puntuación otorgada por el usuario. A continuación, se proporciona un ejemplo de una de las críticas de la colección de datos.

```

• Positivas y negativas = 2625 criticas
<review author="Torbe" title="Los Increibles" rank="5" maxRank="5" source="muchocine">
  <summary>¿Por qué las películas animadas son mucho mejores que las normales?</summary>
  <body>Es que no la cagan en ninguna, todas las pelis que hacen son cada vez mejores. Incluso
para darle apretones y muerdos. Es despues, cuando se echa to el pelo pa los laos cuando esta para ha
ania tienen los yanquis en engrandecer y festejar continuamente la mediocridad. En referencia a ir a l
</review>
    
```

Figura 3 Ejemplo de una crítica de cine del corpus Muchocine.

3.3 Análisis y etiquetado de los datos

Para clasificar las críticas del corpus Muchocine, es necesario separar la crítica de la clase a la que pertenece. Para ello, se extrae el contenido de las etiquetas <body> y <summary>, ya que los resultados presentados por (Martínez-Cámara et al., 2011) muestran que combinar la crítica y su resumen mejora significativamente la clasificación. Además, se extrae el valor numérico de la etiqueta <rank> para identificar la clase a la que pertenece. Si el valor es 1 o 2, se clasifica como crítica negativa (clase 0.0); si es 3 o 4, se clasifica como crítica positiva (clase 1.0). Al finalizar el proceso, se obtiene una tabla de datos separados y etiquetados. A continuación, en la Tabla 2, se muestra la separación entre el texto de la crítica y su correspondiente clase.

Tabla 2 Resultado del etiquetado de datos.

Row No.	clase	mi-texto
1214	1.0	Una cinta independiente, brillante, que se debate entre la inteligente hilaridad y el mas honesto drama
1215	1.0	Bergman en estado puro. Duro a la par que sensible, críptico y desgarrador.El día que me dispuse a vi
1216	1.0	Por fin vemos un Almodóvar que retrata escenas de la vida real, sin adornos, sin extrañezas. Una loa a
1217	0.0	Silicona, esteroides, pactos demoníacos y otras basuras habituales son la base que sustentan esta a
1218	0.0	Una comedia entretenida y poca cosa más para ver una tarde de domingoZoom nos cuenta la historia

3.4 Construcción y pesado de características

Antes de construir las características, se lleva a cabo un preprocesamiento en el texto. Para llevar a cabo este proceso, se utilizó el componente "Tokenize" del software RapidMiner, el cual ofrece funcionalidades de tokenización y eliminación de información innecesaria. Mediante este componente, se eliminaron de forma automática los saltos de línea, números, acentos, signos de puntuación y símbolos, dejando únicamente las letras. Este componente contiene una documentación detallada sobre su funcionalidad y características dentro del mismo software. A continuación, se presenta un ejemplo de un texto antes de aplicar el preprocesamiento.

"May, ¿Quieres ser mi amigo?" es una de esas películas que nos recuerdan que el terror no siempre lleva garras de acero en una mano o una máscara en la cara. El terror y la locura se encuentran mucho más cerca de nosotros, de la realidad, de nuestra pacífica y hasta a veces aburrida monotonía. May funciona bajo el método: la bestia duerme dentro de nosotros. En cada uno de nosotros hay un posible psicópata y nuestra vecina la del segundo puede ocultar un oscuro pasado o una doble vida. Para ello, Lucky McKee, nos narra efectivamente una historia presuntamente cotidiana (que sin embargo, engancha desde el principio) sobre una joven muy parecida a esas otras tantas que pululan a nuestro alrededor. La rarita de la clase, la niña tímida de la tienda de comestibles, esa extraña hermana de nuestro amigo, nuestra prima la del pueblo. Es por esto que nos montamos en un tren de cercanías y no en un tren de alta velocidad con escenas más vistas que las reposiciones de "El Príncipe de Bell Air" y adolescentes lelos e insoportables con superávit de hormonas y déficit de neuronas.

Figura 4 Ejemplo de un fragmento de texto del corpus Muchocine, sin aplicar preprocesamiento.

Después de aplicar preprocesamiento al texto, el resultado sería el siguiente.

May Quieres ser mi amigo es una de esas películas que nos recuerdan que el terror no siempre lleva garras de acero en una mano o una máscara en la cara El terror y la locura se encuentran mucho más cerca de nosotros de la realidad de nuestra pacífica y hasta a veces aburrida monotonía May funciona bajo el método la bestia duerme dentro de nosotros En cada uno de nosotros hay un posible psicópata y nuestra vecina la del segundo puede ocultar un oscuro pasado o una doble vida Para ello Lucky McKee nos narra efectivamente una historia presuntamente cotidiana que sin embargo engancha desde el principio sobre una joven muy parecida a esas otras tantas que pululan a nuestro alrededor La rarita de la clase la niña tímida de la tienda de comestibles esa extraña hermana de nuestro amigo nuestra prima la del pueblo Es por esto que nos montamos en un tren de cercanías y no en un tren de alta velocidad con escenas más vistas que las reposiciones de El Príncipe de Bell Air y adolescentes lelos e insoportables con superávit de hormonas y déficit de neuronas

Figura 5 Ejemplo del texto después de aplicar preprocesamiento.

Después del preprocesamiento del texto, se aplica la técnica de *stemming* utilizando el componente "*Stemming Porter*" en RapidMiner. Este componente implementa el algoritmo de *stemming* Porter, desarrollado por Martin Porter en 1980, que es uno de los algoritmos de *stemming* más utilizados en el procesamiento de lenguaje natural.

El algoritmo de *stemming* Porter busca reducir las palabras a su forma base o raíz eliminando las terminaciones y sufijos. Para lograr esto utiliza una serie de reglas y transformaciones fonéticas. Por ejemplo, palabras como "recomendable", "recomendamos", "recomendar" y "recomendación" se reducirían a la forma base "recomend" (Dubiau & Ale, 2013).

La elección de este método se debe a su amplia utilización y su eficacia en la reducción de palabras a su forma raíz, lo cual puede ayudar a simplificar el texto y eliminar redundancias. Sin embargo, es importante

tener en cuenta que el *stemming* puede generar algunas imprecisiones o pérdida de información, ya que se eliminan las terminaciones que pueden contener significado adicional. A pesar de ello, existen casos en los que el *stemming* puede no ser la mejor opción, ya que puede afectar la comprensión y precisión de las palabras en ciertos contextos como en el trabajo de (Martínez-Cámara et al., 2011).

En la etapa de experimentación, se evalúa el impacto del algoritmo de *stemming* Porter en los resultados, esto permite una selección adecuada de técnicas de procesamiento de texto que se ajusten a las necesidades y características específicas de las críticas de cine en el método propuesto de análisis de sentimientos.

Una vez realizado el preprocesamiento del texto, se procede a la construcción de las características para generar la bolsa de n-gramas. En el método propuesto, se generan n-gramas de diferentes tamaños para posteriormente combinarlos en una sola bolsa de palabras. Para llevar a cabo esta generación de n-gramas, se utiliza el componente "Generate n-gramas (Terms)" en el software RapidMiner.

En el proceso de generación de n-gramas, se configura el parámetro "max_length" del componente en 3, lo que indica que se generarán todos los n-gramas hasta un tamaño máximo de 3. Esto nos permite obtener una combinación de unigramas, bigramas y trigramas. Es importante mencionar que utilizamos RapidMiner para realizar esta tarea debido a su capacidad para procesar y generar n-gramas de manera eficiente.

Con base en la combinación de n-gramas descrita en el estudio de (Sari et al., 2018), se exploran cuatro combinaciones diferentes: unigramas + bigramas, unigramas + trigramas, bigramas + trigramas y unigramas + bigramas + trigramas. El objetivo de esta exploración es determinar si la

combinación de n-gramas mejora los resultados de clasificación de sentimientos e identificar cuál de estas combinaciones ofrece los mejores resultados.

En la representación gráfica de los n-gramas, los n-gramas de tamaño 1 se muestran en color negro, los de tamaño 2 en rojo y los de tamaño 3 en verde. Esta visualización permite tener una mejor comprensión de cómo se generan y combinan los diferentes n-gramas en nuestra metodología.

May May_quieres May_quieres_ser quieres quieres_ser quieres_ser_mi ser ser_mi
ser_mi_amigo mi mi_amigo mi_amigo_May amigo amigo_May amigo_May_Quieres May
May_Quieres May_Quieres_ser Quieres Quieres_ser Quieres_ser_mi ser ser_mi
ser_mi_amigo mi mi_amigo mi_amigo_es amigo amigo_es amigo_es_una es es_una
es_una_de una una_de una_de_esas de de_esas de_esas_películas esas esas_películas
esas_películas_que películas películas_que películas_que_nos que que_nos
que_nos_recuerdan nos nos_recuerdan nos_recuerdan_que recuerdan recuerdan_que
recuerdan_que_el que que_el que_el_terror el el_terror el_terror_no terror terror_no
terror_no_siempre no no_siempre no_siempre_lleva siempre siempre_lleva
siempre_lleva_garras lleva lleva_garras lleva_garras_de garras garras_de garras_de_acero
de de_acero de_acero_en acero acero_en acero_en_una en en_una en_una_mano una
una_mano una_mano_o mano mano_o mano_o_una

Figura 6 Palabras, bigramas y trigramas de palabra obtenidos a partir del texto.

A continuación, se lleva a cabo la etapa de pesado de términos, también conocida como "pesado de características".

3.5 Pesado de características

La etapa de pesado de características tiene como objetivo asignar un valor de peso a cada n-grama, en función de su importancia en el texto. En el método propuesto, se utiliza la técnica *TF-IDF* (Term Frequency-Inverse Document Frequency) para asignar estos valores de peso, basándonos en los resultados presentados por Martínez-Cámara et al. (2011).

TF-IDF es una técnica ampliamente utilizada en el procesamiento de lenguaje natural que permite evaluar la importancia de un término dentro de un documento en relación con una colección de documentos, esta técnica se compone de dos componentes principales:

1. Frecuencia del término (TF - Term Frequency): mide la frecuencia con la que aparece un término específico en un documento. Cuantas más veces aparezca un término en un documento, mayor será su valor de frecuencia del término.
2. Frecuencia inversa del documento (IDF - Inverse Document Frequency): mide la importancia de un término en toda la colección de documentos. Se calcula como el logaritmo del cociente entre el número total de documentos y el número de documentos en los que aparece el término. Si un término aparece en pocos documentos, se considera más relevante.

Al combinar estos dos componentes, *TF-IDF* asigna un valor de peso a cada n-grama (entre 0 y 1), reflejando su relevancia en el contexto del análisis de sentimientos en críticas de cine. La propuesta tiene como base los resultados presentados por Martínez-Cámara et al. (2011), que demostraron que *TF-IDF* ofrece los mejores resultados en comparación con otras técnicas de pesado de términos utilizadas en el análisis de texto. Al utilizar *TF-IDF*, destacamos los n-gramas más relevantes y distintivos, lo que nos permite capturar mejor la información importante para la clasificación de sentimientos.

La siguiente tabla muestra el valor de *TF-IDF* de algunos n-gramas, además muestra cómo se combinan los diferentes tamaños de n-gramas en una sola bolsa de palabras, dicho de otra forma, en un solo vector.

Tabla 3 Ejemplo de pesado de términos.

Texto No.	A	A_destacar	A_diferencia	A_estas	A_estas_al turas
1	0.008	0	0	0	0
2	0.24	0	0	0.046	0.47
3	0.13	0	0	0.053	0.054
4	0	0	0	0	0
5	0	0	0	0	0
6	0	0	0	0	0
7	0.006	0	0	0	0
8	0	0	0	0	0
9	0	0	0	0	0
10	0.014	0	0	0	0

La Tabla 3 presenta una organización de los textos y las características generadas. La columna 1 indica el número de cada texto, mientras que las columnas restantes representan diferentes características, desde palabras individuales (columna 2), hasta bigramas (columnas 3, 4 y 5) y trigramas de palabras (columna 6). Dado que el número de características generadas puede ser elevado, es necesario llevar a cabo una reducción de características para evitar un proceso de aprendizaje demasiado lento, a esta etapa también se le conoce como selección de características.

3.6 Selección de características

En esta etapa, se seleccionan las características que serán utilizadas por el clasificador. Debido a la falta de estudios previos sobre la combinación de características para críticas de cine en español, no existe una referencia clara sobre cuántas características deben ser seleccionadas. Para el método

propuesto, se ponen a prueba diferentes cantidades de características, para conocer cual logra obtener mejores resultados.

3.7 Entrenamiento del modelo de clasificación

En este trabajo, se utilizaron dos clasificadores: SVM y Redes Neuronales. SVM obtuvo los mejores resultados en el idioma español (Martínez-Cámara et al., 2011). Por otro lado, los métodos basados en Redes Neuronales ocupan los primeros lugares en el ranking de métodos para el análisis de sentimientos en idioma inglés, según los resultados obtenidos en (*Papers with Code*, 2023). Un ejemplo destacado es el trabajo de (Yang et al., 2020), cuyo método basado en Redes Neuronales se posiciona en el primer lugar del ranking en cuanto a los resultados de críticas de cine en el corpus IMDB.

3.7.1 Clasificador

En primer lugar, se elige el clasificador a utilizar en este caso se prueban dos opciones: SVM y Redes Neuronales. Dentro de SVM, se utiliza la variante SVM Lineal, la cual ha obtenido buenos resultados en el trabajo de (Martínez-Cámara et al., 2011). Este clasificador se basa en la capacidad de separar dos clases de datos (Vapnik, 1998); además, se emplean otras variantes como SVM Polinomial y SVM *Primal*. Por otro lado, las Redes Neuronales han demostrado tener un gran potencial en la clasificación de textos, según (Lai et al., 2015), por lo que se evaluarán también junto con la combinación de n-gramas.

3.8 Evaluación del método

En este trabajo, se realiza una selección cuidadosa de las métricas utilizadas para evaluar la eficacia de los clasificadores en nuestro método de análisis de sentimientos. La elección se basa en la necesidad de obtener

una visión integral y precisa del desempeño del método propuesto. En particular, se realiza una comparación con dos trabajos previos (Cruz et al., 2008; Martínez-Cámara et al., 2011) que utilizaron el corpus Muchocine como referencia. Ambos trabajos anteriores se enfocaron principalmente en la métrica de precisión para evaluar la calidad de los clasificadores. Sin embargo, se toma en consideración la importancia de otras métricas para obtener una evaluación más completa. Por ello, tomamos como referencia el trabajo de (Martínez-Cámara et al., 2011) que incluye las métricas de *Recall* y *F-measure*, ampliamente utilizadas en la recuperación de información (Sebastiani, 2002). Sin embargo, este trabajo propone ampliar la evaluación con dos métricas adicionales: *Accuracy* y *Kappa*. A continuación se explica en que consiste cada métrica.

- *Accuracy*: Es una métrica que mide la proporción de predicciones correctas realizadas por el modelo en relación con el total de muestras. Se calcula dividiendo el número de predicciones correctas entre el número total de muestras. Es una medida general de qué tan bien el modelo puede clasificar correctamente las muestras en general.
- *Precision*: Se refiere a la proporción de verdaderos positivos con relación a la suma de los verdaderos positivos y los falsos positivos. Se centra en cuán precisas son las predicciones positivas del modelo. Una alta precisión indica que el modelo tiene una baja tasa de falsos positivos y es capaz de identificar correctamente las muestras positivas.
- *Recall*: Se refiere a la proporción de verdaderos positivos con relación a la suma de los verdaderos positivos y los falsos negativos. Mide la capacidad del modelo para encontrar y capturar correctamente todas las muestras positivas. Un alto *Recall* indica que el modelo tiene una

baja tasa de falsos negativos y es capaz de identificar correctamente la mayoría de las muestras positivas.

- *Accuracy*: proporciona una medida general del clasificador que correctamente tanto los casos positivos como los casos negativos. Permite determinar el porcentaje de instancias clasificadas correctamente en comparación con el total de instancias.
- *Kappa*: ofrece una evaluación de la concordancia entre las predicciones del clasificador y las etiquetas reales, teniendo en cuenta la posibilidad de que las coincidencias se produzcan por azar. *Kappa* es particularmente útil cuando existe un desequilibrio en la distribución de clases o cuando las tasas de acierto pueden ser influidas por el azar. Proporciona una medida más robusta y precisa del rendimiento del clasificador.

Para facilitar la comprensión de las métricas de evaluación, se proporciona una imagen a continuación.

		Respuesta del método	
		+	⊖
Respuesta correcta	+	Verdadero positivo	Falso negativo
	⊖	Falso positivo	Verdadero negativo

Figura 7 Matriz de confusión (Macarena, 2016)

La matriz de confusión es una valiosa herramienta para evaluar la precisión de un clasificador. En este caso, se están considerando dos clases: críticas positivas y críticas negativas. Cada elemento de la matriz representa una combinación específica de respuestas correctas e incorrectas del clasificador. Para una comprensión completa de cómo utilizar la matriz de confusión, es importante considerar las descripciones proporcionadas por expertos en el tema, como (Macarena, 2016), (Wachsmuth, 2015)).

- 1- Verdadero positivo (VP): significa que el método acertó con la predicción, ya que el modelo clasificó el elemento como positivo, cuando en realidad era positivo.
- 2- Falso negativo (FN): significa que el método no acertó con la predicción, ya que el modelo clasificó el elemento como negativo, cuando en realidad era positivo.
- 3- Falso positivo (FP): significa que el método no acertó con la predicción, ya que el modelo clasificó el elemento como positivo, cuando en realidad era negativo.
- 4- Verdadero negativo (VN): significa que el método acertó con la predicción, ya que el modelo clasificó el elemento como negativo, cuando en realidad era negativo.

Establecidos los cuatro elementos anteriores las fórmulas para las métricas con las que se evalúa el método propuesto son las siguientes:

Ecuaciones de las métricas a utilizar

$$\mathbf{Precision} = \frac{VP}{VP+FP}$$

$$\mathbf{Recall} = \frac{VP}{VP+FN}$$

$$\mathbf{Accuracy} = \frac{VP+VN}{VP+VN+FP+FN}$$

$$\mathbf{F-measure} = \frac{2VP}{2VP+FP+FN}$$

$$\mathbf{Kappa} = \frac{po-pe}{1-pe}$$

Donde:

po = *Accuracy*

$$\mathbf{pe} = \frac{[(vp+fp)(vp+fn)+(fn+vn)(fp+vn)]}{vp+fp+vn+fn}$$



Capítulo 4

Experimentos y discusión

En este capítulo se evalúa el rendimiento del método propuesto, presentándose de manera detallada cada experimento realizado. Se ponen a prueba diferentes combinaciones de tamaños de n-gramas, número de características, uso de *stemming* y variantes de SVM y Redes Neuronales. Para medir la efectividad del método se utilizarán cinco métricas diferentes: *Precision*, *Accuracy*, *Recall*, *F-measure* y *Kappa*.

Para garantizar la precisión y la replicabilidad de los resultados, se aplica el método *10-cross validation*, como se hizo en los experimentos del estado del arte (Cruz et al., 2008; Martínez-Cámara et al., 2011). Este método proporciona resultados más fiables y permite un análisis más detallado de los mismos. El *10-cross validation* consiste en dividir el conjunto de datos en 10 partes iguales, utilizando cada una de ellas como conjunto de prueba una vez, mientras que las otras nueve partes se utilizan como conjunto de entrenamiento. Esto se repite 10 veces para obtener una medida más confiable del rendimiento del método, promediando los resultados obtenidos en cada iteración.

En cuanto al análisis de datos, se emplean tres herramientas: RapidMiner, PowerBI y Python. RapidMiner y PowerBI son reconocidos como excelentes software para el análisis de datos, ciencia de datos y aprendizaje automático (Idoine, 2019; Woodyer, 2019). Por su parte, Python es

ampliamente utilizado en la comunidad de inteligencia artificial y ciencia de datos (Puget, 2017).

A continuación, se describe cada uno de los hiperparámetros que utilizan SVM y Redes Neuronales, detallando en qué consiste y mostrando los valores asignados a cada uno de ellos en la tabla correspondiente:

Tabla 4 Hiperparámetros de SVM.

Hiperparámetro	Descripción	Valor
<i>C</i>	El parámetro C se define como un parámetro de regularización. Controla el equilibrio entre la maximización del margen y la minimización del término de error de entrenamiento.	<i>1</i>
<i>Epsilon</i>	Es el valor que determina cuánto error estamos dispuestos a tolerar en la clasificación. Un <i>epsilon</i> más pequeño hace que el modelo sea más estricto en la clasificación, mientras que un valor mayor permite cierto grado de error.	0.001
<i>Bias</i>	Es un ajuste adicional que permite al modelo tener un margen de maniobra para clasificar correctamente los datos. Ayuda a mejorar la precisión del modelo al tener en cuenta factores adicionales.	<i>True</i>

Tabla 5 Hyperparámetros de la Red neuronal

Hiperparámetro	Descripción	Valor
<i>Activation function</i>	Determina la salida de una neurona y su capacidad para activarse o desactivarse.	<i>Sigmoid</i>
<i>Neurons</i>	Unidades individuales en una red neuronal que procesan información y generan salidas.	25000
<i>Loss</i>	Mide que tan bien está funcionando la red neuronal al comparar las salidas predichas con las salidas reales.	<i>Binary crossentropy</i>
<i>Optimizer</i>	Ajusta los pesos y sesgos de la red neuronal durante el entrenamiento para minimizar la función de pérdida.	<i>RMSprop</i>
<i>Learning rate</i>	Controla la magnitud de los ajustes realizados a los pesos y sesgos durante el entrenamiento.	0.001
<i>Epsilon</i>	Un pequeño valor utilizado para evitar divisiones por cero.	1.00E-08
<i>Epochs</i>	Representa el número de veces que todo el conjunto de entrenamiento se pasa a través de la red neuronal durante el entrenamiento.	1000
<i>Batch size</i>	El número de ejemplos de entrenamiento utilizados en una iteración del algoritmo de entrenamiento.	12

4.1 Modelo de n-gramas simple

En esta sección, se evalúa el rendimiento de tres modelos diferentes de n-gramas: unigramas, bigramas y trigramas, cuando se utiliza SVM como clasificador. Lo anterior con el fin de conocer que tamaño de n-grama da los mejores resultados y compararlo con el modelo propuesto de n-gramas combinados.

Tabla 6 Comparación de rendimiento entre diferentes tamaños de n-gramas de palabra utilizando SVM.

Característica	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-measure</i>	<i>Kappa</i>
Unigramas	88.21	88.19	87.5	87.84	0.764
Bigramas	80.61	79.84	80.47	80.16	0.612
Trigramas	79.09	78.29	78.91	78.6	0.582

En la Tabla 6, se presenta la comparación del rendimiento de los tamaños de n-gramas (unigramas, bigramas y trigramas) utilizando SVM como clasificador. De acuerdo con los resultados, el tamaño de unigramas Y SVM presenta una mayor *Accuracy* con un 88.21%, seguido de bigramas con 80.61% y trigramas con 79.09%.

En cuanto a la métrica precisión, unigramas también presenta el mayor valor con 88.19%, seguido de bigramas con 79.84% y trigramas con 78.29%.

En cuanto al *Recall*, unigramas presenta un valor superior a bigramas y trigramas con 87.5% en comparación con 80.47% y 78.91% respectivamente.

De igual manera en *F-measure*, unigramas nuevamente presenta un valor superior con 87.84%, seguido de bigramas con 80.16% y trigramas con 78.6%.

Por último, en la métrica *Kappa*, bigramas presenta un valor mayor que trigramas con 0.612% en comparación con 0.582%. Sin embargo, ambos valores son inferiores al valor presentado por unigramas con 0.764%.

En conclusión, los resultados indican que los unigramas presentan un mejor rendimiento en las diferentes métricas evaluadas. Esto se debe a que los unigramas representan las palabras individuales en el texto, lo que permite capturar la información más básica y directa presente en las críticas de cine. Al considerar únicamente las palabras aisladas, se evita la complejidad y ambigüedad que pueden surgir al considerar combinaciones de palabras más largas.

Además, los unigramas tienen la ventaja de ser más frecuentes y tener una mayor cobertura en el corpus de texto. Al haber más instancias de unigramas en comparación con bigramas o trigramas. Es decir, el clasificador tiene más información disponible para realizar la clasificación de sentimientos.

Tabla 7 Comparación de rendimiento entre diferentes tamaños de *n*-gramas de palabra utilizando Redes Neuronales.

Característica	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-measure</i>	<i>Kappa</i>
Unigramas	87.48	87.39	87.4	86.84	0.744
Bigramas	79.43	77.79	78.44	78.14	0.604
Trigramas	78.07	77.23	77.08	76.9	0.565

Como se observa en la tabla 7, al utilizar Redes Neuronales los resultados obtenidos son consistentes con los del experimento anterior. Los unigramas continúan mostrando un rendimiento superior en términos de *Accuracy*, *Precision*, *Recall*, *F-measure* y *Kappa*. Por otro lado, los tamaños

de n-grama bigrama y trigrana presentan resultados inferiores en comparación con el unigrana.

Al analizar la métrica *Kappa*, se evidencia que tanto el unigrana como el bigrama exhiben resultados similares, mientras que el trigrana muestra el valor más bajo. Estos hallazgos sugieren que existe una coherencia en el comportamiento de los diferentes tamaños de n-grama en relación con las métricas evaluadas.

En general, estos resultados indican que el unigrana es el tamaño de n-grama más apropiado para este experimento, incluso cuando se emplean Redes Neuronales como clasificador. Esta consistencia en los patrones de desempeño confirma la preferencia por el uso del unigrana como la opción más favorable en este contexto.

4.2 Combinación de n-gramas

En este experimento se combinan diferentes tamaños de n-gramas para crear una sola bolsa de n-gramas. En la Tabla 8 se muestran los resultados de las diferentes combinaciones posibles entre unigramas, bigramas y trigramas.

Tabla 8 Comparación de la combinación de las bolsas n-gramas.

Característica	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-measure</i>	<i>Kappa</i>
Unigramas+bigramas	88.97	88.98	88.28	88.63	0.779
Unigramas+trigramas	88.97	85.6	83.59	84.58	0.703
Bigramas+trigramas	85.17	79.23	80.47	79.84	0.604
Unigramas+bigramas +trigramas	90.87	91.94	89.06	90.48	0.817

De acuerdo a los resultados, se puede ver que la mejor combinación fue unigramas + bigramas + trigramas, con una *Accuracy* del 90.87%, una *Precision* del 91.94%, un *Recall* del 89.06%, un *F-measure* del 90.48% y *Kappa* 0.817%. Esto indica que la combinación de tres tamaños de n-gramas, proporciona un mejor rendimiento en comparación a la combinación de solo dos tamaños.

En segundo lugar, se encuentra la combinación de unigramas + bigramas, con una *Accuracy* del 88.97%, una precisión del 88.98%, un *Recall* del 88.28% y un *F-measure* del 88.63%. Esto sugiere que la combinación de unigramas y bigramas también proporciona un buen rendimiento, aunque no tan alto como la combinación de tres tamaños.

Por último, las combinaciones de unigramas y trigramas y bigramas y trigramas presentaron resultados cercanos. Sin embargo, estos resultados aún están muy por debajo de la mejor combinación de n-gramas.

En conclusión, los resultados sugieren que la combinación de tres tamaños de n-gramas proporciona el mejor rendimiento en comparación con la combinación de solo dos tamaños. Esto puede ser útil para mejorar la precisión de modelos de lenguaje en el futuro.

4.3 Stemming

Con la finalidad de conocer si la técnica de *stemming* mejora los resultados, en la Tabla 9 se comparan los resultados obtenidos con y sin la técnica de *stemming* en el preprocesamiento. Se utiliza el modelo de n-gramas combinado que obtuvo los mejores resultados en un experimento previo, es decir, la combinación de unigramas + bigramas + trigramas de palabra. La Tabla 9 muestra unos resultados superiores cuando se utiliza *stemming* en la etapa de preprocesamiento en comparación de no utilizarlo.

Tabla 9 Aplicación de stemming en la etapa de preprocesamiento.

Característica	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-measure</i>	<i>Kappa</i>
Sin stemming	90.87	91.94	89.06	90.48	0.817
Con stemming	87.45	88.62	85.16	86.85	0.749

4.4 Número de características

En el presente estudio, se llevó a cabo un análisis del número de características (n-gramas) que se utilizaran para el método propuesto. Con el objetivo de evaluar el impacto del número de características en el rendimiento del modelo, se realizaron varios experimentos utilizando distintas cantidades de n-gramas.

En la Tabla 10, se muestran los resultados obtenidos al variar el número de características empleadas en el análisis:

Tabla 10 Número de características utilizadas

Número de característica	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>	<i>F-measure</i>	<i>Kappa</i>
20,000	90.54	91.77	88.87	90.32	0.809
22,000	90.68	91.84	88.96	90.39	0.812
24,000	90.79	91.90	89.02	90.45	0.815
25,000	90.87	91.94	89.06	90.48	0.817
26,000	90.83	91.92	89.04	90.46	0.816
28,000	90.71	91.86	88.96	90.4	0.813
30,000	90.63	91.81	88.91	90.36	0.81
32,000	90.56	91.78	88.87	90.33	0.809
34,000	90.48	91.75	88.82	90.29	0.807

En particular, se observa que al emplear 25,000 características se obtuvieron los mejores resultados. Los resultados obtenidos respaldan la importancia de considerar un número adecuado de atributos para lograr un análisis de sentimientos preciso y confiable en el contexto de las críticas de cine en español. El valor óptimo de 25,000 características demostró ser el más efectivo en este estudio, brindando resultados superiores en términos de *Accuracy*, *Precision*, *Recall*, *F-measure* y *Kappa*.

4.5 Variantes del clasificador

El clasificador SVM es un algoritmo que ha demostrado ser muy efectivo en muchas aplicaciones de aprendizaje automático, por lo que no es de extrañar que haya sido ampliamente utilizado en trabajos previos del estado del arte (Cruz et al., 2008; Martínez-Cámara et al., 2011). Sin embargo, hasta ahora, no se ha evaluado el rendimiento de las variantes de SVM en el corpus actual. Por lo tanto, en los experimentos siguientes, se evalúan las variantes de SVM (Lineal, *Primal* y Polinomial) para determinar si alguna de ellas tiene un rendimiento superior al clásico SVM Lineal. La combinación de unigramas + bigramas + trigramas, que fue la que obtuvo los mejores resultados en un experimento anterior, se utiliza como base de comparación.

Tabla 11 Comparación del rendimiento de las variantes de SVM

Clasificador	<i>Accuracy</i>	Precisión	<i>Recall</i>	<i>F-measure</i>	<i>Kappa</i>
SVM <i>Primal</i>	90.87	91.94	89.06	90.48	0.817
SVM Lineal	89.73	91.06	87.50	89.24	0.794
SVM Polinomial	84.79	87.29	80.47	83.74	0.695

De acuerdo con los resultados mostrados en la Tabla 11, el clasificador SVM *Primal* obtuvo un *Accuracy* del 90.87%, lo que indica que fue capaz de clasificar correctamente el 90.87% de las instancias evaluadas. La precisión de este clasificador fue del 91.94%, lo que muestra la proporción de predicciones positivas que fueron realmente correctas. El *Recall*, que representa la proporción de instancias positivas correctamente identificadas, fue del 89.06%. La *F-measure*, que combina la precisión y el *Recall*, alcanzó un valor de 90.48%. Por último, el valor de *Kappa* fue de 0.817, lo que indica un buen nivel de acuerdo entre las predicciones y las clases reales. Estos resultados son superiores a los obtenidos por los otros dos clasificadores. Por otro lado, el SVM Lineal obtuvo un segundo lugar en las 5 métricas evaluadas, ya que sus resultados fueron superiores a los obtenidos por SVM Polinomial pero inferiores a los obtenidos por SVM *Primal*. Por último, los resultados de SVM Polinomial fueron inferiores a los de las otras dos variantes de SVM.

En conclusión, la Tabla 11 muestra que el SVM *Primal* tiene los mejores resultados de los tres clasificadores, seguido por el SVM Lineal y el SVM Polinomial. Estos resultados pueden ayudar a tomar una decisión sobre cuál variante de SVM utilizar en futuros experimentos o proyectos.

4.5 Discusión

Se observa un empate en términos de rendimiento entre SVM *Primal* y dual. A pesar de que se ha logrado una precisión del 91.94%, se observa un incremento de tan solo un 0.88% en comparación con el SVM Lineal estándar. Este hallazgo indica que la clave para mejorar el rendimiento no radica en la variante del clasificador, sino en el proceso de preprocesamiento. La combinación de unigramas, bigramas y trigramas en el preprocesamiento logra un aumento del 3.75% en comparación con no combinar los n-gramas. Además, el uso de *stemming* no produce ninguna

mejora en los resultados y también retrasó el proceso de clasificación debido al procesamiento adicional. Por esta razón, no se recomienda utilizar *stemming* con este método.

Después de completar la experimentación, se presenta el mejor resultado obtenido, que se logró combinando unigramas, bigramas, trigramas y utilizando SVM *Primal*. Este resultado supera los estándares actuales en el campo del análisis de sentimientos en críticas de cine en español.

Tabla 12 Comparación del método propuesto con el estado del arte.

Autores	Clasificador	Precisión	Recall	F-measure
Método propuesto	SVM <i>Primal</i>	91.94	89.06	90.48
Martinez [6]	SVM Lineal	87.73	87.69	87.71
Martinez [6]	Naive Bayes	84.08	84.01	84.04
Cruz [5]	<i>Sov1</i> - S	77.50	N/A	N/A
Cruz [5]	<i>Sov2</i> - NS	69.05	N/A	N/A

*N/A No se cuenta con la información, ya que no se evaluó con esta métrica

La Tabla 12 presenta una comparación del método propuesto con el estado del arte. Los resultados indican que el método propuesto, al utilizar un clasificador SVM *Primal*, muestra un incremento en la precisión en comparación con el SVM Lineal estándar. Se ha observado una mejora significativa, alcanzando una precisión del 91.94%, un *Recall* del 89.06% y una medida F del 90.48%. Estos resultados superan a los obtenidos por otros autores, como Martinez [6] quien utiliza un clasificador SVM Lineal y obtiene una precisión del 87.73%, un *Recall* del 87.69% y una medida F del 87.71%. Además, también supera al resultado obtenido por Martinez [6] utilizando un clasificador Naive Bayes, con una precisión del 84.08%, un

Recall del 84.01% y una medida F del 84.04%. Finalmente, el método propuesto supera a los resultados obtenidos por Cruz [5] utilizando los clasificadores *S0v1 - S* y *S0v2 - NS* con una *Precision* del 77.50% y 69.05% respectivamente, aunque no se cuenta con la información de *Recall* y *F-measure* para estos resultados.

Se concluye que el método propuesto es una solución más efectiva que los métodos del estado del arte en este corpus, y que la combinación de n-gramas es la clave para lograr un mejor rendimiento. La elección del clasificador se vuelve secundaria cuando se utiliza una combinación de características en el modelo de representación.

Los resultados demuestran una mejoría en todas las métricas evaluadas en comparación con los métodos del estado del arte, logrando una precisión del 91.94%. Sin embargo, aún hay margen de mejora para alcanzar la meta de una precisión del 99.9%. Este trabajo es un punto de partida para futuros estudios que busquen optimizar la precisión y aplicarla en la vida real.



Capítulo 5

Conclusiones y aportaciones

En este capítulo se presentan las conclusiones, aportaciones y trabajos futuros que surgen de esta investigación.

5.1 Conclusiones

A continuación, se resumen los hallazgos más relevantes de la aplicación del método propuesto en el corpus Muchocine:

- 1- El método propuesto logra un rendimiento superior en términos de precisión, *Recall* y *F-measure* en comparación con los métodos del estado del arte descritos por (Cruz et al., 2008; Martínez-Cámara et al., 2011). Se logra un rendimiento de 91.94 siendo superior al de los métodos presentados en la literatura revisada para el corpus Muchocine el cual tenían como máximo 87.73% de precisión .
- 2- El clasificador SVM *Primal* obtiene la mejor precisión, es un 0.88% más preciso que SVM Lineal, 6.08% en comparación con SMV Polinomial y supera por 4.21% la mejor precisión que se tenía en estado del arte.
- 3- Los resultados obtenidos por SVM y Redes Neuronales son muy similares, siendo superior SVM por aproximadamente 1%.
- 4- Combinar n-gramas mejora los resultados obtenidos en comparación con no combinarlos para la tarea de análisis de sentimientos, lo que abre camino a futuras investigaciones que combinen distintos tipos de características.

- 5- En este trabajo se incluye la exploración de la técnica de *stemming*, a pesar de que se habían reportado resultados negativos en el estado del arte. Sin embargo, es importante resaltar que nuestro método difiere en su enfoque y características particulares en el análisis de sentimientos. Aunque se había informado que el *stemming* afectó negativamente los resultados en otros métodos, decidimos experimentar con esta técnica en el método propuesto para evaluar su efecto.
- 6- Finalmente, no se encontró una gran diferencia entre los resultados obtenidos por cada métrica, debido a la similar cantidad de críticas positivas y negativas evaluadas.

5.2 Aportaciones

- 1- Se demuestra la importancia de la combinación de n-gramas como modelo de representación, lo cual abre nuevas posibilidades de investigación en el futuro para combinar distintos tipos de características.
- 2- Se propone un nuevo método para el análisis de sentimientos en críticas de cine que logra un rendimiento de 91.94 siendo superior al de los métodos presentados en la literatura revisada para el corpus Muchocine el cual tenían como máximo 87.73% de precisión.
- 3- Se realiza una descripción detallada del proceso de investigación, ilustrando cómo se llevaron a cabo los experimentos y se obtuvieron los resultados.
- 4- Los resultados muestran que el clasificador SVM obtuvo una precisión superior en relación con las Redes Neuronales. SVM alcanzó una precisión del 91.94%, mientras que las Redes Neuronales obtuvieron 91.06%. Dicha diferencia puede ser

relevante en la aplicación práctica del análisis de sentimientos, destacando la efectividad del SVM como clasificador en este dominio.

- 5- Se demuestra que la técnica de *stemming* empeora los resultados en el análisis de sentimientos en críticas de cine para el idioma español utilizando SVM y Redes Neuronales, lo cual es una aportación en la literatura sobre el tema.

5.3 Trabajos futuros

- 1- Investigar la eficacia de Redes Neuronales recurrentes como clasificador.
- 2- Evaluar el uso de la red Bert en la clasificación de sentimientos.
- 3- Mejorar el rendimiento de SVM a través de la optimización con algoritmos genéticos.
- 4- Evaluar el rendimiento de word2vec en la clasificación de sentimientos.
- 5- Aplicar y evaluar el método propuesto en otros corpus para ampliar su alcance y validez.



Referencias

- Abambres, M., Marcy, M., & Doz, G. (2018). Potential of Neural Networks for Structural Damage Localization. *SSRN Electronic Journal*.
<https://doi.org/10.2139/ssrn.3368672>
- Abdulla, N. A., Ahmed, N. A., Shehab, M. A., & Al-Ayyoub, M. (2013). Arabic sentiment analysis: Lexicon-based and corpus-based. *2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, 1–6.
<https://doi.org/10.1109/AEECT.2013.6716448>
- Balahur, A., & Turchi, M. (2012). *Multilingual Sentiment Analysis using Machine Translation?*
- Brooke, J., Tofiloski, M., & Taboada, M. (2009). *Cross-Linguistic Sentiment Analysis: From English to Spanish*.
- Can, E. F., Ezen-Can, A., & Can, F. (2018). *Multilingual Sentiment Analysis: An RNN-Based Framework for Limited Data*. Undefined.

/paper/Multilingual-Sentiment-Analysis%3A-An-RNN-Based-for-Can-Ezen-Can/08b536c7d4175a5143fa89b1c6bf7ab5702a39df

Cervantes, I. (2021). *EL ESPAÑOL EN EL MUNDO 2021. Anuario del Instituto Cervantes.*

Crow, D. (2022, julio 6). *The Failure of Lightyear: Does Disney Have a Streaming Problem?* Den of Geek.
<https://www.denofgeek.com/movies/failure-of-lightyear-disney-have-streaming-problem/>

Cruz, F. L., Troyano, J. A., Enriquez, F., & Ortega, J. (2008). Clasificación de documentos basada en la opinión: Experimentos con un corpus de críticas de cine en español. *Procesamiento del Lenguaje Natural*, 41(0), Article 0.
<http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/2551>

de Albornoz Cuadrado, J. C. (2011). *Un Modelo Lingüístico-Semántico Basado en Emociones para la—Tenis De Mesa Y Mesa De Ping-Pong* [Universidad Complutense de Madrid].
<http://doczz.es/doc/205325/un-modelo-ling%C3%BC%C3%ADstico-sem%C3%A1ntico-basado-en-emociones-para-la>

Dubiau, L., & Ale, J. M. (2013). *Análisis de Sentimientos sobre un Corpus en Español: Experimentación con un Caso de Estudio.* 12.

Estrada, C. A., & Fabela, J. L. T. (2022). *Nuevo método para atribución de autoría en muestras de entrenamiento balanceadas y desbalanceadas del corpus C10.*

Forbes Staff. (2016, agosto 15). *Adaptarse a la era digital, clave para pequeñas y grandes empresas.* Forbes México. <https://www.forbes.com.mx/el-futuro-de-las-empresas-esta-en-internet/>

Ge, J., Alonso-Vazquez, M., & Gretzel, U. (2017). *Sentiment Analysis: A Review.*

Google. (2020). *Utiliza las GPU para modelos de entrenamiento en la nube.* Google Cloud. <https://cloud.google.com/ai-platform/training/docs/using-gpus?hl=es-419>

Idoine, C. (2019). *Magic Quadrant for Data Science and Machine Learning Platforms.* Gartner. <https://www.gartner.com/en/documents/3899464/magic-quadrant-for-data-science-and-machine-learning-pla>

Khanna, D., Sahu, R., Baths, V., & Deshpande, B. (2015). Comparative Study of Classification Techniques (SVM, Logistic Regression and Neural Networks) to Predict the Prevalence of Heart Disease. *International Journal of Machine Learning and Computing*, 5(5), 414–419. <https://doi.org/10.7763/IJMLC.2015.V5.544>

Kim, S., Koh, Y., Cha, J., & Lee, S. (2015). Effects of social media on firm value for U.S. restaurant companies. *International Journal of*

- Hospitality Management*, 49, 40–46.
<https://doi.org/10.1016/j.ijhm.2015.05.006>
- Lai, S., Xu, L., Liu, K., & Zhao, J. (2015). Recurrent convolutional neural networks for text classification. *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2267–2273.
- Liu, B. (2012). Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*.
<https://doi.org/10.2200/S00416ED1V01Y201204HLT016>
- Luo, X., Zhang, J., & Duan, W. (2013). Social Media and Firm Equity Value. *Information Systems Research*, 24(1), 146–163. JSTOR.
- Macarena. (2016, mayo 16). *Un acercamiento a los modelos de clasificación. Inteligencia Analítica*. <https://inteligencia-analitica.com/acercamiento-modelos-clasificacion/>
- Martínez-Cámara, E., Martín-Valdivia, M. T., & Ureña-López, L. A. (2011). Opinion Classification Techniques Applied to a Spanish Corpus. En R. Muñoz, A. Montoyo, & E. Métais (Eds.), *Natural Language Processing and Information Systems* (Vol. 6716, pp. 169–176). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-22327-3_17
- Miranda, C. H., & Guzmán, J. (2017). Una Revisión Sobre el Análisis de Sentimientos en Español. *Tecciencia*, 12(22), 35–48.
<https://doi.org/10.18180/tecciencia.2017.22.5>

- Mirończuk, M. M., & Protasiewicz, J. (2018). A recent overview of the state-of-the-art elements of text classification. *Expert Systems with Applications*, 106, 36–54.
<https://doi.org/10.1016/j.eswa.2018.03.058>
- Mohammad, S. M. (2017). Challenges in Sentiment Analysis. En E. Cambria, D. Das, S. Bandyopadhyay, & A. Feraco (Eds.), *A Practical Guide to Sentiment Analysis* (Vol. 5, pp. 61–83). Springer International Publishing. https://doi.org/10.1007/978-3-319-55394-8_4
- Papers with Code*. (2023). <https://paperswithcode.com/>
- Plakias, S., & Stamatatos, E. (2008). Tensor Space Models for Authorship Identification. En J. Darzentas, G. A. Vouros, S. Vosinakis, & A. Arnellos (Eds.), *Artificial Intelligence: Theories, Models and Applications* (Vol. 5138, pp. 239–249). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-87881-0_22
- Puget, J.-F. (2017). The Most Popular Language For Machine Learning and Data Science Is *KDnuggets*. <https://www.kdnuggets.com/the-most-popular-language-for-machine-learning-and-data-science-is.html/>
- Sari, Y., Stevenson, M., & Vlachos, A. (2018). Topic or Style? Exploring the Most Useful Features for Authorship Attribution. *Proceedings of the 27th International Conference on Computational Linguistics*, 343–353. <http://www.aclweb.org/anthology/C18-1029>

Sebastiani, F. (2002). Machine learning in automated text categorization. *ACM Computing Surveys*, 34(1), 1–47.
<https://doi.org/10.1145/505282.505283>

Tamrakar, C., Pyo, T.-H., & Gruca, T. (2018). *Social Media Sentiment and Firm Value*.

The Sprout Social Index, Edition VIII: Turned Off. (2016). Sprout Social.
<https://sproutsocial.com/insights/data/q3-2016/>

Top Brand Values Report for Global 500 2020 | Brandirectory. (2020).
<https://brandirectory.com/rankings/global/>

Tripathi, P., Vishwakarma, S. K., & Lala, A. (2015). Sentiment Analysis of English Tweets Using Rapid Miner. *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*.
<https://doi.org/10.1109/CICN.2015.137>

Vapnik, V. (1998). *Statistical learning theory* new york. NY: Wiley, 1(2), 3.

Vijayan, V. K., Bindu, K. R., & Parameswaran, L. (2017). A comprehensive study of text classification algorithms. *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 1109–1113. <https://doi.org/10.1109/ICACCI.2017.8125990>

Wachsmuth, H. (2015). *Text Analysis Pipelines: Towards Ad-hoc Large-Scale Text Mining*. Springer International Publishing.
<https://www.springer.com/gp/book/9783319257402>

Woodyer, A. (2019). *Invest Implications: ‘Magic Quadrant for Analytics and Business Intelligence Platforms’*. Gartner.

<https://www.gartner.com/en/documents/3902167/invest-implications-magic-quadrant-for-analytics-and-bus>

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., & Le, Q. V. (2020). *XLNet: Generalized Autoregressive Pretraining for Language Understanding* (arXiv:1906.08237; Versión 2). arXiv. <http://arxiv.org/abs/1906.08237>