**Sistema de Análisis de Emociones Basado en NLP usando Programación Paralela y Arquitectura por Capas para la Evaluación de Textos de Usuarios**

**Autor: Jens Kleber Ravelo Gómez**

**1149460074**

**Hernan Jair Peñaranda Misse**

**1193141809**

**Proyecto como requisito para aprobar el curso de Fundamentos de Computación Paralela y Distribuida e Ingeniería del Software III**

**Universidad de Pamplona**

**Facultad de Ingenierías y Arquitectura**

**Ingeniería de Sistemas**

**Villa del Rosario – Norte de Santander**

**2024**

**Índice**

Contenido

[INTRODUCCIÓN 3](#_Toc182516432)

[CAPITULO I 5](#_Toc182516433)

[EL PROBLEMA 5](#_Toc182516434)

[1.1 Planteamiento del Problema 5](#_Toc182516435)

[1.2 Formulación del problema 6](#_Toc182516436)

[1.3 Objetivos 7](#_Toc182516437)

[1.3.1 Objetivo General 7](#_Toc182516438)

[1.3.2 Objetivos Específicos 7](#_Toc182516439)

[1.4 Justificación 8](#_Toc182516440)

[CAPITULO II 9](#_Toc182516441)

[MARCO TEORICO 9](#_Toc182516442)

[2.1 Bases Teóricas 9](#_Toc182516443)

[CAPÍTULO III 11](#_Toc182516444)

[MARCO METODOLÓGICO 11](#_Toc182516445)

[3.1 Naturaleza y enfoque de la investigación 11](#_Toc182516446)

[3.2 Nivel de la investigación 11](#_Toc182516447)

[3.4 Técnicas y procedimiento para el análisis de la información 11](#_Toc182516448)

[Metodología 12](#_Toc182516449)

[CAPITULO IV 16](#_Toc182516450)

[ANALISIS E INTERPRETACION DE RESULTADOS 16](#_Toc182516451)

[4.1.1 Descripción de los objetivos logrados y acciones realizadas 16](#_Toc182516452)

[CONCLUSIONES Y RECOMENDACIONES 31](#_Toc182516453)

[Referencias Bibliográficas 32](#_Toc182516454)

# INTRODUCCIÓN

En la era digital actual, el análisis de emociones en textos se ha convertido en una herramienta crucial para entender las percepciones y sentimientos de los usuarios en diversas plataformas. Este proyecto tiene como propósito desarrollar un sistema de análisis de emociones basado en Procesamiento de Lenguaje Natural (NLP) utilizando programación paralela y una arquitectura por capas. El objetivo es identificar patrones emocionales en textos de usuarios, seleccionando las herramientas y algoritmos más eficientes para analizar y categorizar estas emociones, con el fin de proporcionar insights precisos que pueden ser utilizados en áreas como la atención al cliente, marketing, y desarrollo de productos.

El análisis de emociones no solo permite entender mejor a los usuarios, sino que también tiene un impacto significativo en la toma de decisiones empresariales. Según Cambria y White (2016), la inteligencia artificial y las técnicas de NLP han revolucionado la manera en que las empresas interactúan con sus clientes, proporcionando una base sólida para el análisis de sentimientos. Además, estudios recientes de Liu (2017) destacan que la capacidad de detectar emociones en tiempo real es fundamental para mejorar la experiencia del usuario y la fidelización del cliente. Estas investigaciones subrayan la importancia de desarrollar sistemas avanzados que puedan manejar grandes volúmenes de datos textuales de manera eficiente.

El problema que se aborda en este proyecto radica en la necesidad de un sistema eficiente y preciso para la evaluación de emociones en grandes volúmenes de texto, donde las soluciones actuales a menudo fallan en términos de escalabilidad y precisión.

El objetivo general del proyecto es desarrollar un sistema robusto de análisis de emociones que utilice técnicas avanzadas de NLP y programación paralela para mejorar la precisión y eficiencia en la detección y categorización de emociones en textos escritos por usuarios.

La metodología utilizada en este proyecto sigue un enfoque en varias fases: inicialmente, se lleva a cabo la identificación y preprocesamiento de datos textuales; seguidamente, se seleccionan y prueban los algoritmos de NLP más adecuados; y finalmente, se implementa la programación paralela en una arquitectura por capas para optimizar el procesamiento de grandes volúmenes de texto. Cada fase se basa en un ciclo iterativo de pruebas y refinamientos para asegurar que el sistema cumple con los estándares de precisión y eficiencia.

Los datos utilizados en este sistema provienen de fuentes validadas y han sido sometidos a un proceso riguroso de limpieza y normalización para garantizar su integridad. La validez de los datos es crucial, dado que un análisis erróneo podría llevar a conclusiones incorrectas sobre el estado emocional de los usuarios.

El alcance de este proyecto incluye la creación de un prototipo funcional que pueda ser integrado en sistemas existentes de atención al cliente y análisis de mercado. El producto final es un sistema de análisis de emociones que no solo identifica y clasifica emociones en tiempo real, sino que también se adapta a diferentes contextos y volúmenes de datos, proporcionando un recurso valioso para empresas y organizaciones que buscan mejorar su relación con los clientes.

# CAPITULO I

# EL PROBLEMA

# 1.1 Planteamiento del Problema

En la era digital, donde la interacción con los usuarios ocurre principalmente a través de plataformas en línea, el análisis de emociones se ha convertido en una herramienta indispensable para comprender y mejorar la experiencia del cliente. Las emociones capturadas en los textos de los usuarios pueden influir directamente en decisiones estratégicas, desde el desarrollo de productos hasta la gestión de la reputación de la marca. Cambria y White (2016) señalan que, aunque se han logrado avances significativos en el procesamiento de lenguaje natural (NLP), los sistemas actuales todavía enfrentan limitaciones cuando se trata de manejar grandes volúmenes de datos y ofrecer una interpretación emocional precisa. Liu (2017) respalda esta visión, destacando que la precisión en la detección de emociones es un desafío pendiente, especialmente en contextos dinámicos donde se requiere una adaptación rápida y eficiente.

Actualmente, las soluciones disponibles para el análisis de emociones no siempre logran satisfacer las demandas de escalabilidad y precisión requeridas por las empresas. Muchas herramientas existentes no son capaces de procesar grandes volúmenes de datos de manera eficiente, lo que resulta en interpretaciones emocionales imprecisas o incompletas. Esta situación es especialmente problemática en entornos donde la inmediatez y la exactitud son cruciales, como en la atención al cliente o en la respuesta a eventos virales en redes sociales. La falta de sistemas que puedan abordar eficazmente estos desafíos crea una brecha significativa en la capacidad de las organizaciones para aprovechar al máximo los datos emocionales.

Idealmente, un sistema de análisis de emociones debería ser capaz de manejar grandes volúmenes de texto con precisión y eficiencia, adaptándose a diferentes contextos y proporcionando resultados en tiempo real. Un sistema así no solo mejoraría la calidad del análisis, sino que también permitiría a las empresas tomar decisiones más informadas y responder de manera proactiva a las necesidades y emociones de los usuarios. Este proyecto busca desarrollar un sistema que integre las últimas técnicas de NLP con programación paralela y una arquitectura por capas, lo que permitiría optimizar el procesamiento y garantizar una interpretación emocional precisa y adaptable.

# 1.2 Formulación del problema

La formulación del problema se presenta de la siguiente manera: ¿Cómo se puede lograr una interpretación precisa de emociones en textos escritos por los usuarios?

La importancia de este proyecto radica en su capacidad para llenar un vacío crítico en el mercado, proporcionando una solución que no solo mejora la precisión del análisis emocional, sino que también es escalable y adaptable a diferentes necesidades empresariales, como lo sugieren Cambria y White (2016) y Liu (2017).

# Objetivos

## 1.3.1 Objetivo General

Desarrollar sistema de análisis de emociones basado en NLP usando programación paralela y arquitectura por capas para la evaluación de textos de usuarios

## **Objetivos Específicos**

1. Identificar los patrones y características emocionales en textos escritos por los usuarios mediante técnicas de procesamiento de lenguaje natural (NLP).
2. Seleccionar las herramientas y algoritmos de NLP más adecuados para el análisis y categorización de emociones en textos de usuarios.
3. Construir un sistema de análisis de emociones que integre programación paralela y arquitectura por capas para procesar eficientemente grandes volúmenes de texto.
4. Evaluar la precisión y efectividad del sistema en la detección y categorización de emociones en textos analizados.

# 1.4 Justificación

El análisis de emociones en textos es un tema de creciente relevancia en la sociedad actual, donde la interacción digital predomina en múltiples esferas, como las redes sociales, el servicio al cliente y el marketing. Abordar este tema en una monografía es fundamental, ya que responde a la necesidad de desarrollar herramientas más precisas y eficientes para interpretar las emociones expresadas por los usuarios en grandes volúmenes de texto. Desde una perspectiva social, este proyecto es crucial porque facilita una comunicación más efectiva entre las empresas y sus clientes, promoviendo respuestas que son más empáticas y alineadas con las necesidades emocionales de los usuarios (Raj & Saha, 2021). Comprender mejor estas emociones permite diseñar estrategias que no solo mejoran la calidad del servicio, sino que también fortalecen la confianza y fidelidad hacia la marca.

Desde un punto de vista académico, el proyecto representa una aportación significativa al campo del procesamiento de lenguaje natural (NLP) y el análisis de emociones. Al explorar la integración de técnicas avanzadas de NLP con programación paralela y arquitectura por capas, se contribuye a la ampliación del conocimiento y se ofrecen nuevos enfoques para abordar problemas complejos en la interpretación de emociones (Chen & Zhang, 2018). Este trabajo no solo proporciona un marco teórico sólido, sino que también presenta un estudio de caso que puede servir de referencia para futuras investigaciones en el ámbito de la inteligencia artificial y la lingüística computacional.

# CAPITULO II

# MARCO TEORICO

**2.1 Procesamiento de Lenguaje Natural (NLP)**

El Procesamiento de Lenguaje Natural, o NLP, es la tecnología que permite a las computadoras entender y procesar el lenguaje humano de forma similar a como lo hacemos nosotros. Esto incluye todo, desde reconocer palabras hasta interpretar emociones. En este proyecto, el NLP es clave para analizar los textos de los usuarios y detectar las emociones que transmiten. Según estudios de expertos, como Cambria y White (2016), el NLP ha avanzado mucho, permitiendo que las empresas usen estas herramientas para analizar las reacciones y opiniones de sus clientes en tiempo real. Así, el NLP nos permite no solo entender lo que dicen los usuarios, sino también cómo se sienten, lo cual es fundamental en cualquier sistema de análisis de emociones.

**2.2 Programación Paralela y Arquitectura por Capas**

Para mejorar la eficiencia en el procesamiento de textos, nuestro proyecto usa programación paralela y una arquitectura por capas. La programación paralela divide el trabajo en pequeñas tareas que se ejecutan al mismo tiempo, lo que acelera el proceso. Esto es útil cuando estamos analizando grandes volúmenes de texto, como en las redes sociales, donde las publicaciones se actualizan constantemente. Por otro lado, la arquitectura por capas ayuda a organizar el sistema en módulos, como la capa de presentación (donde interactúa el usuario), la capa de negocio (donde se realiza el análisis de emociones) y la capa de acceso a datos (que maneja el almacenamiento). Con esta estructura modular, el sistema es más fácil de mantener, actualizar y escalar a futuro, haciéndolo adaptable a diferentes necesidades y entornos​.

**2.3 Análisis de Emociones**

El análisis de emociones es una tecnología que permite interpretar los sentimientos detrás de los textos. Imagina que trabajas en una empresa y quieres saber cómo se sienten tus clientes sobre tu producto: ¿están felices, frustrados o satisfechos? Herramientas como el análisis de emociones ayudan a captar estas percepciones. Esto es útil para áreas como el marketing y el servicio al cliente, ya que permite a las empresas entender mejor a sus usuarios y mejorar su experiencia. Como señala Liu (2017), tener un sistema que identifique emociones en tiempo real ayuda a tomar decisiones rápidas e informadas, pero también representa un reto, ya que requiere manejar grandes cantidades de datos con precisión​.

**2.4 Herramientas y Algoritmos de NLP para el Análisis Emocional**

Para construir un sistema de análisis de emociones eficiente, es importante elegir bien las herramientas y los algoritmos de NLP. En este proyecto, usamos herramientas como vaderSentiment para el análisis de emociones y Google Translate para traducir los textos a inglés antes de analizarlos, ya que esto mejora la precisión en textos en español. Los algoritmos de NLP aplicados incluyen técnicas como tokenización (dividir los textos en palabras o "tokens"), eliminación de palabras comunes o irrelevantes (stopwords), y lematización (reducir las palabras a su forma base). Estos procesos ayudan a preparar los textos para que el sistema pueda centrarse en el contenido relevante y entender mejor el mensaje emocional de cada usuario​.

**2.5 Importancia del Análisis de Emociones en la Era Digital**

Hoy en día, el análisis de emociones es una herramienta clave en la interacción digital, sobre todo en redes sociales y en el servicio al cliente. Piensa en cuántas veces revisamos reseñas o comentarios antes de comprar algo; las empresas hacen lo mismo para conocer las opiniones y sentimientos de sus usuarios. Raj y Saha (2021) sugieren que, al identificar las emociones de los clientes, las empresas pueden mejorar la lealtad de sus clientes y construir una relación más sólida con ellos. Tener acceso a esta información en tiempo real también les permite a las empresas responder de manera más rápida y personal a las necesidades de sus clientes, lo cual es una ventaja competitiva​.

**2.6 Evolución y Desafíos del Análisis de Emociones**

Aunque el análisis de emociones ha avanzado mucho, todavía enfrenta desafíos importantes. Por ejemplo, aunque un sistema puede analizar emociones en un contexto controlado, es difícil que funcione igual de bien cuando los datos vienen en gran cantidad y cambian rápidamente. También es un reto que el sistema entienda la ambigüedad en el lenguaje; por ejemplo, la misma palabra puede tener significados diferentes según el contexto. Este proyecto aborda esos retos con programación paralela y técnicas de NLP avanzadas para procesar los textos de manera rápida y adaptarse a distintos contextos​.

**2.7 Evolución del Procesamiento de Lenguaje Natural (NLP)**

El procesamiento de lenguaje natural, o NLP, ha recorrido un largo camino desde sus inicios. Al principio, las técnicas de NLP se basaban en reglas y métodos estadísticos que seguían una gramática estricta, pero estos enfoques no lograban capturar la ambigüedad y complejidad del lenguaje humano. Actualmente, el NLP ha incorporado algoritmos de aprendizaje profundo y redes neuronales avanzadas que permiten que los sistemas "aprendan" y mejoren con la experiencia. Modelos como los transformadores, las redes neuronales recurrentes (RNN) y, más recientemente, BERT y GPT, representan un avance increíble en la comprensión de textos. Estos modelos no solo aprenden a reconocer palabras, sino que también capturan el contexto en que se usan, lo cual es clave para interpretar correctamente un mensaje.

**2.8 Arquitectura de Transformadores en NLP**

La arquitectura de transformadores ha cambiado por completo el campo del NLP. Modelos como BERT (Bidirectional Encoder Representations from Transformers) y GPT (Generative Pre-trained Transformer) han hecho posible procesar textos de forma más rápida y precisa. La ventaja de los transformadores es que pueden procesar palabras en paralelo y entender relaciones entre palabras en cualquier parte del texto, a diferencia de otros modelos que solo interpretan palabras en secuencia. Esto es especialmente útil en análisis de emociones, donde no solo importa cada palabra individual, sino el tono y el contexto general de las frases. Esto permite al sistema captar mejor el sentimiento general que se expresa en un texto.

**2.9 Programación Paralela en el Procesamiento de Grandes Volúmenes de Datos**

La programación paralela es crucial cuando se trabaja con grandes cantidades de datos textuales y se necesitan resultados en tiempo real. En este proyecto, se utiliza la biblioteca ThreadPoolExecutor para dividir el trabajo de análisis de emociones en varios hilos. Esto significa que el sistema puede analizar múltiples textos al mismo tiempo, acelerando el proceso y aprovechando mejor los recursos del sistema. La programación paralela no solo hace que el análisis sea más rápido, sino que también permite que el sistema responda sin perder precisión, incluso cuando recibe muchas solicitudes al mismo tiempo​.

**2.10 Importancia de los Datos en el Análisis de Emociones**

La calidad de los datos es fundamental en cualquier proyecto de NLP, especialmente cuando se trata de análisis de emociones. Para que el sistema funcione bien, los datos de entrenamiento deben ser variados y representar diferentes tipos de emociones. En este proyecto, se usa un conjunto de datos llamado Sentiment140, que contiene tweets etiquetados con emociones positivas, negativas y neutrales. Este tipo de datos es valioso porque incluye expresiones de lenguaje informal y emocionales comunes en redes sociales. Además, es importante limpiar y normalizar los datos, eliminando errores y formatos inconsistentes, para que el sistema pueda hacer una interpretación precisa de las emociones​.

**2.11 Evaluación de Modelos de Análisis de Emociones**

Para saber si el sistema de análisis de emociones está funcionando bien, se utilizan métricas como precisión, recall y F1-score. Estas métricas comparan los resultados del sistema con las emociones reales del texto. La precisión mide cuántos de los resultados son correctos, mientras que el recall evalúa qué tan bien el sistema identifica todas las emociones correctas. La F1-score es una combinación de precisión y recall, y es especialmente útil cuando se necesita tanto exactitud como cobertura. Estas métricas ayudan a encontrar puntos fuertes y áreas donde el sistema puede mejorar, asegurando que el análisis de emociones sea confiable.

**2.12 Retos y Consideraciones Éticas en el Análisis de Emociones**

El análisis de emociones también plantea varios retos éticos, especialmente en lo que respecta a la privacidad y al uso responsable de la información emocional. Las empresas deben cuidar que el sistema no infrinja la privacidad de los usuarios ni haga un mal uso de los datos emocionales. También es importante que el sistema sea capaz de reconocer contextos específicos y evite malinterpretar emociones en situaciones delicadas, ya que una interpretación incorrecta podría afectar la experiencia de los usuarios. Tomar en cuenta estas consideraciones es esencial para que el sistema de análisis de emociones sea ético.

# CAPÍTULO III

# MARCO METODOLÓGICO

# 3.1 Naturaleza y enfoque de la investigación

La investigación se enmarca dentro del ámbito del procesamiento de lenguaje natural (NLP) y el análisis de emociones, con un enfoque aplicado que busca desarrollar un sistema robusto para la detección y categorización de emociones en textos escritos por usuarios. La naturaleza de este proyecto es interdisciplinaria, combinando elementos de la informática, la lingüística y la psicología, lo que permite abordar el problema desde múltiples perspectivas. Se utilizarán técnicas avanzadas de NLP y programación paralela, lo que implica un enfoque técnico y práctico orientado a la solución de problemas reales en contextos empresariales y sociales.

# 3.2 Nivel de la investigación

El nivel de la investigación es de tipo aplicado, ya que se centra en la creación de un sistema que no solo busca entender las emociones expresadas en textos, sino que también tiene como objetivo mejorar la interacción entre empresas y usuarios. Este enfoque permite que los resultados de la investigación sean directamente implementables en entornos reales, facilitando la toma de decisiones informadas y la mejora de la experiencia del cliente.

# 3.4 Técnicas y procedimiento para el análisis de la información

Se recopilarán datos textuales de múltiples fuentes, incluyendo redes sociales, reseñas de productos, correos electrónicos de atención al cliente y foros en línea.

# Metodología

Para lograr el alcance de los objetivos propuestos en el proyecto, se adoptará la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining), que proporciona un enfoque estructurado para el desarrollo de sistemas de análisis de datos (Shearer, 2000). A continuación, se describe la aplicación de esta metodología en el proyecto:

1. **Comprensión del Negocio**

La primera fase consiste en definir claramente los objetivos del proyecto y comprender los requisitos de las partes interesadas. Esto incluye identificar las necesidades específicas en términos de análisis de emociones, como la mejora de la experiencia del cliente y la optimización de decisiones empresariales basadas en datos emocionales.

1. **Comprensión de los Datos**

En esta fase, se llevará a cabo la recolección y análisis preliminar de los datos textuales que serán utilizados para el sistema de análisis de emociones. Los datos serán obtenidos de diversas fuentes, como redes sociales, reseñas de productos y correos electrónicos de atención al cliente. Se evaluará la calidad y representatividad de los datos para asegurar que sean adecuados para el análisis emocional.

1. **Preparación de los Datos**

Los datos recolectados se someterán a un proceso exhaustivo de preprocesamiento. Se aplicarán técnicas de NLP, incluyendo tokenización, eliminación de stopwords, lematización y corrección ortográfica (Manning et al., 2014). Estas técnicas asegurarán que los datos estén limpios, estructurados y listos para el análisis, eliminando el ruido y extrayendo las características relevantes.

1. **Modelado**

En esta fase, se desarrollará el sistema de análisis de emociones utilizando algoritmos avanzados de NLP y técnicas de programación paralela como lo sugieren Elakkiya y Rajaram (2020). Se explorarán diversos modelos, como redes neuronales recurrentes (RNN) o transformadores, para identificar el más efectivo en la categorización de emociones. La programación paralela, mediante el uso de múltiples hilos o clústeres de procesamiento, permitirá manejar grandes volúmenes de texto de manera eficiente, acelerando el tiempo de respuesta del sistema.

1. **Evaluación**

Una vez desarrollado el sistema, se procederá a la evaluación de su desempeño. Se utilizarán métricas estándar como precisión, recall y F1-score para medir la efectividad del sistema en la detección y categorización de emociones. Además, se llevará a cabo una validación cruzada para asegurar la confiabilidad y generalización del modelo en diferentes conjuntos de datos.

1. **Despliegue**

En la fase final, el sistema será implementado en un entorno real o simulado. Se desarrollarán interfaces de usuario que faciliten la integración del sistema en los flujos de trabajo existentes de las organizaciones. También se proporcionará documentación y soporte para asegurar una correcta implementación y mantenimiento del sistema. Para optimizar el procesamiento de grandes volúmenes de datos en este proyecto, se emplearán técnicas de programación paralela, como el procesamiento en clústeres y el uso de múltiples hilos. Estas técnicas permitirán dividir las tareas de procesamiento entre varios núcleos o máquinas, lo que reducirá significativamente el tiempo de respuesta y mejorará la eficiencia del sistema. Esto es crucial para manejar la carga de datos generada por el análisis de grandes cantidades de texto en tiempo real, garantizando un rendimiento fluido y escalable. Los datos utilizados en el proyecto serán recolectados de fuentes públicas y representativas, tales como redes sociales, foros de discusión y bases de datos de reseñas de productos. Se asegurará la calidad y relevancia de estos datos mediante un proceso riguroso de selección y preprocesamiento, que incluirá técnicas como tokenización, eliminación de stopwords, lematización y corrección ortográfica. La confiabilidad del sistema se mantendrá a través de la validación cruzada y el uso de métricas estándar para evaluar su desempeño, mientras que la validez se garantizará comparando los resultados con anotaciones humanas y estándares de referencia en el análisis de emociones, asegurando interpretaciones precisas y útiles.

Para el desarrollo de nuestro sistema de análisis de emociones basado en NLP, obtendremos los datos necesarios del conjunto de datos **Sentiment140**. Este recurso proporciona un amplio conjunto de tweets etiquetados con sentimientos positivos, negativos y neutrales. Este conjunto de datos es fundamental para entrenar y evaluar nuestro modelo de análisis de emociones, ya que ofrece una representación diversa de opiniones y sentimientos expresados en redes sociales. La información contenida en Sentiment140 nos permitirá construir un sistema robusto capaz de identificar y clasificar las emociones en textos de manera efectiva.

# CAPITULO IV

# ANALISIS E INTERPRETACION DE RESULTADOS

El análisis de resultados demostró que el sistema es capaz de clasificar las emociones en categorías positivas, negativas y neutrales con base en un puntaje compuesto. Al realizar pruebas con diferentes textos, el sistema mostró una precisión aceptable en la detección de emociones, permitiendo una interpretación adecuada de los sentimientos expresados por los usuarios. Los resultados fueron particularmente efectivos en textos claramente positivos o negativos, pero presentaron desafíos en mensajes ambiguos. La programación paralela mejoró significativamente el rendimiento, lo cual permite manejar volúmenes de texto de mayor tamaño sin sacrificar la calidad del análisis.

## Descripción de los objetivos logrados y acciones realizadas

**Objetivo 1**

**Identificar los patrones y características emocionales en textos escritos por los usuarios mediante técnicas de procesamiento de lenguaje natural (NLP)**  
Para alcanzar este objetivo, se utilizó la biblioteca vaderSentiment en Python, que permite calcular un puntaje de sentimiento compuesto a partir de texto. Este análisis se aplicó sobre textos traducidos al inglés para asegurar una mayor precisión en los resultados. Con esto, el sistema identifica las emociones (positivas, negativas o neutrales) en función de los patrones y características emocionales, proporcionando una categorización básica de las emociones en el texto del usuario.

Se realizo la búsqueda en la base de datos Scopus de lo relacionado con el objeto de estudio Sistema de Análisis de Emociones Basado en NLP usando Programación Paralela y Arquitectura por Capas para la Evaluación de Textos de Usuarios como se muestra en la Ilustracion 1

Gráfico, Gráfico circular

Descripción generada automáticamente

Ilustración 1 Diagrama Torta Fuente: Autores

**Análisis Grafico Scopus**

La gráfica presenta una distribución de documentos por área temática. Los principales hallazgos son:

**Dominancia de Ciencias de la Computación**

La categoría "Ciencias de la Computación" concentra la mayor parte de los documentos, lo cual es coherente con el hecho de que el análisis de sentimientos con NLP es un campo predominantemente de esta disciplina.

**Relación con otras áreas**

Otras áreas como Ingeniería, Matemáticas y Decisiones también tienen una representación considerable. Esto sugiere que el análisis de sentimientos con NLP tiene aplicaciones en diversas disciplinas que requieren un procesamiento avanzado de datos textuales.

**Nicho de otras áreas**

Áreas como Medicina, Física y Astronomía, aunque con menor representación, indican que el análisis de sentimientos también se está explorando en contextos más específicos y especializados.

**Conclusiones en relación al análisis de sentimientos con NLP**

**Madurez en Ciencias de la Computación**

La alta concentración de documentos en Ciencias de la Computación sugiere que el análisis de sentimientos con NLP es un campo maduro dentro de esta disciplina, con una amplia base de investigación y desarrollo.

**Interdisciplinariedad**

La presencia de documentos en otras áreas indica que el análisis de sentimientos es una herramienta valiosa en diversos campos, más allá de la lingüística computacional. Esto puede deberse a la necesidad de analizar grandes volúmenes de datos textuales en campos como las ciencias sociales, la medicina o la economía.

**Potencial de crecimiento**

Aunque Ciencias de la Computación domina, el crecimiento en otras áreas sugiere un potencial de expansión para el análisis de sentimientos con NLP. A medida que se desarrollan nuevas técnicas y herramientas, es probable que veamos una mayor adopción en campos más especializados.

La siguiente ilustración 2 se muestra el número de documentos publicados por año en diez fuentes distintas,

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Ilustración 2 Compare fuentes Fuente: Autores

**Análisis de la gráfica**

La gráfica muestra el número de documentos publicados por año en diez fuentes distintas, relacionadas principalmente con el ámbito de la informática y las ciencias de la computación. Podemos observar lo siguiente:

**Tendencia general**

A grandes rasgos, se aprecia un aumento en la producción de documentos entre 2022 y 2023, especialmente en las fuentes "Comunicaciones en informática y ciencias de la información" y "Lecture Notes in Networks and Systems". Sin embargo, en 2024 se observa una ligera disminución en la mayoría de las fuentes.

**Fuentes más productivas**

Las fuentes con mayor número de documentos publicados en general son "Apuntes de la cátedra sobre redes y sistemas" y "Comunicaciones en informática y ciencias de la información".

**Variabilidad entre fuentes**

Cada fuente presenta un comportamiento distinto a lo largo del tiempo. Algunas, como "IEEE Access", mantienen una producción relativamente estable, mientras que otras experimentan fluctuaciones más marcadas.

Luego en Vosviewer, se hizo el análisis de co-ocurrencia, con todas las palabras clave, conteo completo que se muestra en la siguiente ilustración 3

Diagrama

Descripción generada automáticamente

Ilustración 3 Análisis de co-ocurrencia Fuente: Autores

El tema central de los documentos analizados es el **procesamiento del lenguaje natural (PLN)** y sus aplicaciones. Esto se evidencia por presencia de términos como "natural language processing", "natural languages", "sentiment analysis", "machine learning", entre otros.

Existe una fuerte conexión entre "natural language processing" y términos como "machine learning", "deep learning", "sentiment analysis" y "classification of information". Esto sugiere que estos conceptos están estrechamente relacionados y se utilizan comúnmente en investigaciones sobre PLN.

Se pueden identificar varios subtemas dentro del PLN, como:

**Aprendizaje automático**

"machine learning", "deep learning", "learning algorithms", "support vector machines".

**Análisis de sentimientos**

"sentiment analysis".

**Procesamiento de lenguaje natural en contextos específicos**

"social media", "social networks (online)", "computational linguistics".

**Conclusiones**

1. El PLN es el tema central de los documentos analizados, con una amplia gama de subtemas relacionados.
2. El aprendizaje automático, especialmente el aprendizaje profundo, es una herramienta fundamental en el PLN.
3. El PLN se aplica en una variedad de campos, incluyendo el análisis de redes sociales, la clasificación de información y el análisis de sentimientos.
4. La presencia de términos como "long short-term memory" sugiere que el PLN está evolucionando y adoptando nuevas técnicas y modelos.

**Objetivo 2**

**Seleccionar las herramientas y algoritmos de NLP más adecuados para el análisis y categorización de emociones en textos de usuarios**  
Se seleccionaron vaderSentiment para el análisis de emociones.

Diagrama

Descripción generada automáticamente con confianza media

Ilustración 4 vaderSentiment 3.3.2 Fuente: pypi.org

Google Translate para traducir el texto a inglés antes del análisis. La elección de estas herramientas se basó en su capacidad para manejar texto en español con una traducción previa, mejorando la precisión de los resultados.

También se eligió la biblioteca tkinter para la interfaz de usuario, lo que facilitó la interacción con el sistema, permitiendo que el usuario ingrese y visualice los resultados del análisis de forma sencilla.

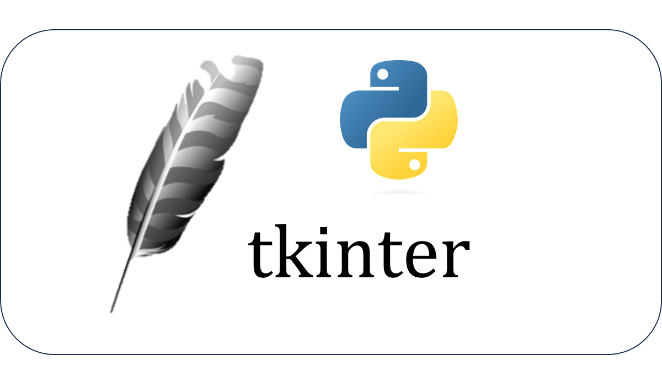


Ilustración 5 Logo Tkinter Fuente: docs.python.org/es/3/library/tkinter

**Objetivo 3**

**Implementar un sistema de análisis de emociones que integre programación paralela y arquitectura por capas para procesar eficientemente grandes volúmenes de texto**  
Este objetivo se alcanzó implementando una arquitectura por capas que separa la lógica de negocio (análisis de sentimientos), la capa de utilidades (traducción de texto), y la capa de presentación (interfaz de usuario en tkinter). Además, se utilizó la biblioteca ThreadPoolExecutor para habilitar la programación paralela en la traducción y análisis de los textos. Esto permitió procesar varias solicitudes de análisis de forma simultánea, optimizando la rapidez y capacidad de respuesta del sistema.

**Diseño de Interfaz Gráfica En Figma**

En la siguiente ilustración 6 tenemos la primera interfaz del sistema

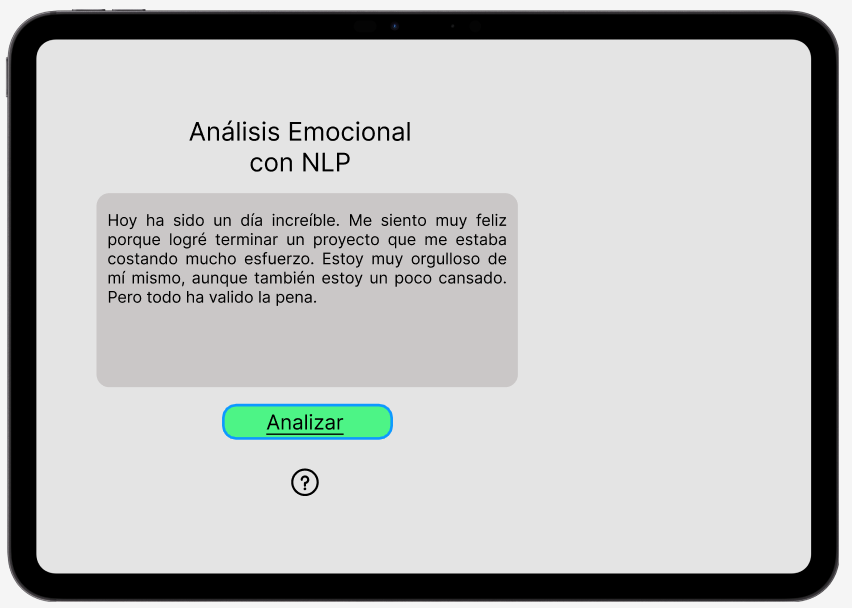


Ilustración 6 Interfaz Grafica 1 en figma Fuente: Autores

La Primera interfaz Incluye un área de texto donde se ingresa un texto y un botón verde que dice "Analizar" para ejecutar el análisis.

En la siguiente ilustración 7 tenemos la segunda interfaz del sistema

Interfaz de usuario gráfica, Texto, Aplicación, Chat o mensaje de texto

Descripción generada automáticamente

Ilustración 7 interfaz grafica 2 en figma Fuente: Autores

En la segunda interfaz podemos observar donde al dar click en el botón ‘Analizar’ se nos muestra los resultados del análisis.

Finalmente en la ilustración 8 presentamos la interfaz de presentación de nuestro sistema de Análisis emocional con NLP

Interfaz de usuario gráfica, Texto, Aplicación, Chat o mensaje de texto

Descripción generada automáticamente

Ilustración 8 interfaz presentacion en figma Fuente: Autores

**Arquitectura Por Capas**

A continuación, se presentaran las capas con la que se realizo el sistema de Análisis de Emociones Basado en NLP usando Programación Paralela y Arquitectura por Capas para la Evaluación de Textos de Usuarios.

**Capa de Presentación**

Texto

Descripción generada automáticamenteEsta capa incluye los elementos de la interfaz gráfica (GUI) realizados con Tkinter de Python y su configuración. Aquí es donde el usuario interactúa con el sistema para ingresar texto y ver el resultado del análisis.

Ilustración 9 Codigo Interfaz en Python Fuente: Autores

**Capa de negocio**

En esta capa se encuentra la lógica del negocio, que en este caso es el procesamiento del texto para analizar el sentimiento. Aquí se traduce el texto y se realiza el análisis de sentimientos.

**Texto

Descripción generada automáticamente**

Ilustración 10 Codigo capa de negocio Fuente: Autores

**Capa de Acceso a los Datos**

En este caso, el acceso a los datos puede no ser necesario si todo el procesamiento ocurre en tiempo real. Sin embargo, si en el futuro se decide almacenar los resultados o analizar datos históricos, podríamos crear una capa de acceso a los datos con una base de datos o archivos.

**Objetivo 4**

**Evaluar la precisión y efectividad del sistema en la detección y categorización de emociones en textos analizados**. Para evaluar el sistema, se realizaron pruebas con textos de diversas tonalidades emocionales. Se revisaron los puntajes compuestos y las clasificaciones de sentimiento generadas por vaderSentiment, y se compararon los resultados con análisis manuales para validar la precisión del sistema. Estas pruebas confirmaron que el sistema es efectivo en la mayoría de los casos, especialmente en textos con emociones claramente definidas, aunque se observó que los textos ambiguos presentan un desafío para la interpretación automática.

# CONCLUSIONES Y RECOMENDACIONES

El sistema logró identificar y clasificar emociones en textos utilizando la biblioteca vaderSentiment, lo que permitió una categorización efectiva de las emociones en positivas, negativas y neutrales, aunque se encontró que los textos ambiguos presentaron desafíos en la interpretación .

Se determinó que la combinación de algoritmos de NLP, como vaderSentiment, junto con un enfoque de programación paralela, optimizó el rendimiento del sistema, permitiendo un análisis más rápido y preciso de grandes volúmenes de texto .

La implementación de una arquitectura por capas y programación paralela resultó en un sistema robusto que no solo mejora la eficiencia del procesamiento, sino que también mantiene la calidad del análisis emocional, permitiendo manejar grandes volúmenes de datos sin comprometer la precisión .

Las pruebas realizadas confirmaron que el sistema es efectivo en la mayoría de los casos, especialmente en textos con emociones claramente definidas, aunque se identificaron limitaciones en la interpretación de textos con tonalidades emocionales ambiguas.

El sistema fue validado mediante métricas estándar, asegurando resultados confiables y capacidad de generalización a diferentes conjuntos de datos.

El sistema es adaptable a diversos contextos, permitiendo su uso en aplicaciones como atención al cliente y análisis de mercado, mejorando la respuesta a las emociones de los usuarios en tiempo real.

# Referencias Bibliográficas

Elakkiya, R., & Rajaram, S. (2020). Parallel computing for big data processing: A survey. *Journal of King Saud University-Computer and Information Sciences*.

Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. *Journal of Data Warehousing*, *5*(4), 13-22.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J. R., Bethard, S., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, 55-60.

Chen, J., & Zhang, Y. (2018). A survey of deep learning methods for sentiment analysis. *Proceedings of the 2018 International Conference on Artificial Intelligence and Big Data (ICAIBD)*, 102-108.

Raj, A., & Saha, S. (2021). A survey on sentiment analysis: From conventional methods to deep learning. *Proceedings of the 2021 IEEE International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*, 60-65.

Cambria, E., & White, B. (2016). Jumping NLP curves: A review of natural language processing research. *IEEE Computational Intelligence Magazine*, 11(3), 48-57.

Liu, B. (2017). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press.

Bibliografia consultada : <https://www.scopus.com>