

Análisis de Sentimientos con Procesamiento del Lenguaje Natural (NLP)



Tecnatura Superior en Ciencia de Datos e Inteligencia Artificial (IA)

Alumnos: Cattaneo Victoria
Chamorro Jesica

Profesor de Práctica III: Walter Ramos

ÍNDICE GENERAL

Introducción	3
Descripción del Proyecto	4
Objetivos generales	4
Desarrollo del proyecto	5
Objetivo del proyecto	5
Recopilar y preparar los datos	5
Preprocesamiento de los datos	5
Herramientas y Tecnologías Utilizadas	6
Futuras Mejoras	7

Introducción

En un mundo en crecimiento con la tecnología, la voz de las personas se elevó gracias al acceso masivo a plataformas de comunicación como redes sociales, blogs, y foros. Las opiniones públicas, ya sean sobre productos, servicios, o temas sociales, tienen un impacto significativo en las decisiones empresariales y gubernamentales.

El análisis de sentimientos es una técnica del campo del Procesamiento de Lenguaje Natural (PLN) que nos permite interpretar y clasificar estas opiniones según la emoción dominante, ya sea satisfacción, insatisfacción, o cualquier otro estado emocional. Esta herramienta no solo ayuda a las empresas a entender mejor a sus clientes, sino que también ofrece a los individuos y comunidades una forma de amplificar sus voces y ser escuchados en temas que les importan.

Este presente proyecto busca implementar un modelo de análisis de sentimientos que pueda predecir las emociones en las opiniones públicas, que pueda ayudar a la toma de decisiones, desde la evaluación de productos, películas, hasta estrategias de marketing.

Esto permitirá no solo reconocer el sentimiento que más predomina en las opiniones sociales, sino también fomentar una empatía y respeto hacia las necesidades de la comunidad.

Descripción del Proyecto

La información no solo se manifiesta a través de imágenes, sino también en las opiniones públicas. En la actualidad, la sociedad ha ganado una voz significativa, y es importante reconocerla mediante la predicción de la emoción dominante en estas opiniones, ya sea de satisfacción o insatisfacción, en relación con la adquisición de un producto o sobre un tema específico. Esto incluye tanto problemas sociales y culturales como preferencias personales, como por ejemplo, opiniones públicas sobre la compra de productos, la calificación de películas, o la adaptación de estrategias de marketing basadas en las opiniones de la comunidad.

El análisis de sentimientos en redes sociales es una técnica que utiliza la inteligencia artificial (IA) para interpretar y clasificar las emociones expresadas en comentarios, publicaciones y menciones en plataformas de compra y venta, redes sociales como X (Twitter), Facebook e Instagram. Dado el vasto volumen de contenido generado en estas plataformas, las empresas y organizaciones utilizan esta herramienta para obtener insights sobre la percepción pública de sus productos, servicios o marcas.

Este análisis implica el procesamiento y la evaluación del texto, donde se identifican sentimientos como positivos, negativos o neutrales. Para llevar a cabo este proceso, se recopilan datos de redes sociales, se preprocesan (limpieza y tokenización del texto), y se aplican modelos de aprendizaje automático o redes neuronales que han sido entrenados para detectar patrones emocionales en el lenguaje.

Objetivos generales

- Lograr determinar si un comentario es positivo o negativo, satisfactorio o insatisfactorio.
- Cuales son las opiniones más redundantes.

Desarrollo del proyecto

Objetivo del proyecto

El análisis de sentimientos es especialmente útil para monitorear la reputación en línea, entender las necesidades del cliente y reaccionar a tiempo ante crisis de imagen. A medida que se obtienen más datos y se mejora el modelo de análisis, las predicciones se vuelven más precisas, lo que permite a las empresas tomar decisiones informadas basadas en el feedback real de los usuarios.

Recopilar y preparar los datos

En este proyecto de análisis de sentimientos en comentarios de Mercado Libre, un paso clave será la obtención de los datos a través de web scraping, extrayendo características como la valoración del producto, la fecha, el comentario y la cantidad de "me gusta".

Una vez obtenidos los datos, el dataset se importará a un entorno de análisis de datos, como Python, utilizando bibliotecas como pandas para su exploración inicial. Esta etapa incluirá la revisión de la estructura del dataset y la limpieza de los datos.

Posteriormente, el dataset será utilizado para realizar predicciones de sentimientos, destacando el sentimiento de cada comentario y su calificación según la predicción. Dado que este proceso es complejo, puede requerir ajustes en el enfoque a lo largo del desarrollo.

Preprocesamiento de los datos

El preprocesamiento de los datos es una etapa crucial en el análisis de sentimientos, ya que asegura que el texto esté en una forma consistente y limpia antes de ser utilizado por los modelos de aprendizaje automático. En esta fase, realizaremos los siguientes pasos:

- **Web Scraping:** Realizaremos web scraping en Mercado Libre para extraer comentarios de uno o dos productos. Extraemos datos como la valoración del producto, la fecha, el comentario y la cantidad de "me gusta" que recibió cada comentario. Comenzaremos identificando patrones en la página web para facilitar la extracción.

- **Limpieza del Texto:** Eliminaremos caracteres especiales, URLs, números y otros elementos no textuales que puedan introducir ruido en los datos. Además, normalizamos el texto convirtiendo todas las palabras a minúsculas para asegurar la uniformidad y evitar que palabras con la misma raíz pero diferente capitalización sean tratadas como diferentes.
- **Tokenización:** Luego, procederemos a dividir el texto en unidades más pequeñas, ya sea en palabras o frases, a través del proceso de tokenización. Esto permite que el modelo de análisis de sentimientos procese el texto de manera más efectiva, reconociendo y analizando cada palabra o grupo de palabras de forma individual.
- **Eliminación de Stopwords:** Finalmente, eliminaremos las stopwords, que son palabras comunes y de poca relevancia semántica, como "el", "de", "y", entre otras. La eliminación de estas palabras ayuda a reducir el ruido en el conjunto de datos y permite que el modelo se concentre en las palabras que realmente aportan valor al análisis del sentimiento.

Estos pasos de preprocesamiento son esenciales para mejorar la calidad del análisis y asegurar que los modelos de IA reciban datos relevantes y libres de ruido, lo que resulta en predicciones más precisas y significativas.

Herramientas y Tecnologías Utilizadas

El análisis se desarrollará utilizando **Jupyter Notebook**, una herramienta interactiva que permite la ejecución de código en Python y facilita la documentación del proceso en un entorno colaborativo. Las principales tecnologías y librerías que se emplearán en este proyecto son:

- **Python:** El lenguaje de programación principal para el desarrollo del proyecto, elegido por su versatilidad y amplia comunidad de soporte en el campo del análisis de datos y el aprendizaje automático.
- **pandas:** Una biblioteca esencial para la manipulación y análisis de estructuras de datos. Se utilizará para la carga, exploración, y limpieza del dataset.
- **numpy:** Librería fundamental para operaciones matemáticas y manejo de matrices, utilizada para el procesamiento numérico de los datos.
- **seaborn y matplotlib:** Estas librerías se emplearán para la visualización de datos, permitiendo crear gráficos y diagramas que faciliten la comprensión y exploración de patrones en los comentarios analizados.
- **scikit-learn (sklearn):** Biblioteca clave en el proyecto, ya que provee herramientas para el preprocesamiento de datos, construcción de modelos de aprendizaje automático y evaluación de su rendimiento.

- **selenium:** Selenium es una herramienta para automatizar navegadores web, útil en pruebas y extracción de datos, controlando acciones como clics o formularios con lenguajes como Python o Java.
- **BeautifulSoup:** Es una librería de Python usada para extraer y analizar datos de páginas web en HTML o XML. Facilita la búsqueda y manipulación de elementos.
- **Otros.**

Futuras Mejoras

El proyecto puede ampliarse en el futuro mediante la incorporación de datos adicionales, el ajuste de modelos o la utilización de técnicas más avanzadas, como transformers o redes neuronales profundas, para mejorar la precisión del análisis.