

P04: Sistema de Detección de Sentimiento en Reseñas de Usuarios

1. Contexto del proyecto

Meta desea comprender mejor la percepción que los usuarios tienen sobre productos, restaurantes y películas. Actualmente, en muchas ocasiones dicho análisis de reseñas se sigue realizando de forma manual o mediante recuentos simples de palabras, dificultando la obtención de conclusiones fiables cuando el volumen de opiniones es elevado.

La organización dispone de conjuntos de datos con reseñas escritas por usuarios y desea explorar el uso de técnicas de aprendizaje automático para identificar automáticamente si una reseña expresa un sentimiento positivo o negativo.

Para mejorar la presentación semántica del texto, se propone utilizar modelos de vectores de palabras y documentos, específicamente el modelo Doc2Vec, junto con un clasificador basado en K-Nearest Neighbors (k-NN), algoritmo de vecinos cercanos.

2. Objetivo del proyecto

Desarrollar un modelo en Python capaz de clasificar las reseñas de los usuarios como positivas o negativas. El texto se representará en la aplicación utilizando Doc2Vec, y se utilizará un algoritmo de k-NN como método de clasificación. Los resultados deben ser entregados de forma comprensible y sencilla para usuarios no expertos en técnicas de inteligencia artificial (gente de marketing digital, publicidad, directivos), mediante una interfaz sencilla y amigable.

El sistema pretende servir como herramienta de apoyo para el análisis de opiniones y la toma de decisiones basada en datos.

3. Interesados del proyecto (stakeholders)

En la siguiente tabla resumen debéis introducir los que pensáis que son los principales interesados en el proyecto y en el producto, y cuáles son sus necesidades y el valor que obtienen del producto.

Interesado	Rol en la institución	¿Qué necesita del sistema?	¿Qué valor obtiene?

4. Necesidades, especificaciones y requerimientos del proyecto.

El director de operaciones de Meta se ha puesto en contacto con vosotros para ofreceros la siguiente información:

Para ofrecer una solución al problema planteado, los datos de entrada serán textos de reseñas en formato textual, almacenados por la propia empresa. La aplicación debe clasificar automáticamente el sentimiento expresado por cada uno de ellos. Para ello, se debe manejar un lenguaje natural realista, incluyendo variaciones en estilo y vocabulario, pudiendo clasificar el mayor amplio espectro de comentarios. En la aplicación final se debe poder introducir un comentario y obtener su clasificación como positivo o negativo. Dicha aplicación debe poder ejecutarse en un entorno estándar sin requerimientos complejos de infraestructura (como un ordenador de con un sistema operativo Windows instalado). Para ello, la dirección sugiere el uso de una pequeña interfaz web utilizando Streamlit.

El proyecto en su totalidad engloba la comprensión y análisis del problema y sus requisitos. Los datos proporcionados deberán ser explorados, tratados y preparados para ser usados en el modelo de clasificación. Se definirá el modelo a utilizar, se entrenará con los datos proporcionados y se evaluarán los resultados del mismo hasta que las predicciones se ajusten con los datos históricos. El modelo entrenado será incorporado en una aplicación sencilla que permita el uso por parte del cliente, e irá acompañado de documentación técnica y funcional.

El acuerdo del proyecto no incluye ni la integración de la aplicación con sistemas reales de comercio electrónico o plataformas de reseñas, ni el despliegue de la misma a gran escala.

Mediante esta descripción, extraed en un listado los requisitos y necesidades de los interesados, así como el alcance del proyecto.

5. Desglose del proyecto en tareas

A continuación, deberéis desglosar el proyecto en un listado de tareas

1. Tarea 1
2. Tarea 2
3. Tarea 3 ...

6. Ciclo de vida de los datos

Realizad una descripción y análisis del ciclo de vida de los datos de vuestro proyecto.

7. Planificación del proyecto

En este apartado deberéis hacer una estimación de la duración de las distintas tareas que habéis desglosado anteriormente, así como de las dependencias entre ellas. Haréis también un listado de los entregables del proyecto. A partir de la tabla de datos, realizaréis un diagrama PERT y un análisis del camino crítico para el proyecto, así como un diagrama Gantt.

Una vez realizado dicho análisis, vais a introducir los recursos (tanto materiales como humanos) y sus restricciones de uso, realizaréis la asignación de recursos a las distintas tareas y volveréis a analizar la duración del proyecto, intentando encontrar los cuellos de botella y los recursos críticos del proyecto.

8. Seguimiento del proyecto

A lo largo del proyecto, utilizaréis una combinación de metodologías ágiles de tipo Scrum y Kanban. Priorizaréis las tareas del proyecto y creareis un gráfico de columnas para representar los estados de las mismas a lo largo de cada Sprint. Realizaréis reuniones de seguimiento todas las semanas durante las sesiones prácticas, y un par de Sprint Reviews después de los entregables parciales. En los documentos entregables de la asignatura debe quedar constancia de las reuniones, los cambios, decisiones y avances a lo largo del proyecto.