



Integrantes:

Liam Marcos Jimenez 2022-0818
Christopher Tolentono 2020-10703
Anthony Volquez Pena 2022-0264

Materia:

Inteligencia Artificial

Maestro:

Ever Miguel Curiel Moran

Asignación:

Tarea: Análisis de Sentimiento de Texto

Fecha:

7/4/2024



Informe del modelo

1. Metodología utilizada

Recolección de datos

Para este proyecto, utilizamos un conjunto de datos de comentarios de Twitter obtenido de la plataforma Kaggle. Este conjunto de datos fue recopilado mediante el uso de la API de Twitter y contiene una amplia variedad de tweets que abordan una variedad de temas y opiniones.

Preprocesamiento de texto

Para preparar los datos para el análisis de sentimientos, se aplicaron los siguientes pasos de preprocesamiento:

1. **Eliminación de caracteres especiales y puntuación:** Se eliminaron los caracteres especiales y signos de puntuación de los comentarios, ya que no contribuyen significativamente al análisis de sentimientos y pueden afectar negativamente la precisión del modelo.
2. **Conversión a minúsculas:** Todos los caracteres de los comentarios se convirtieron a minúsculas para asegurar consistencia en el análisis y simplificar el procesamiento de texto.
3. **Tokenización y eliminación de palabras vacías:** Los comentarios se dividieron en palabras individuales y se eliminaron las palabras vacías, como "el", "la", "de", etc., que no aportan información significativa al análisis de sentimientos.
4. **Lematización para reducir las palabras a su forma base:** Se aplicó lematización a los tokens restantes para reducirlos a su forma base, agrupando palabras con significados similares y mejorando la precisión del análisis de sentimientos al reducir la variabilidad de las palabras.

Una vez que los comentarios fueron preprocesados, se utilizó la técnica TF-IDF (Term Frequency-Inverse Document Frequency) para convertir el texto preprocesado en una representación numérica. TF-IDF asigna pesos a cada palabra en función de su frecuencia en un comentario específico y en el conjunto de datos en su totalidad, lo que permite capturar la importancia relativa de cada palabra en el contexto del análisis de sentimientos.

Entrenamiento del modelo:

Una vez que los datos fueron preprocesados y transformados en una representación numérica adecuada, se procedió al entrenamiento del modelo de análisis de sentimientos. Para este fin, se optó por utilizar un clasificador de regresión logística configurado con el algoritmo 'saga' como método de optimización. Se seleccionó el algoritmo 'saga' debido a su eficiencia y capacidad para manejar problemas grandes de clasificación. Es conocido por ser una elección sólida en términos de rendimiento y escalabilidad.

2. Resultados obtenidos

El modelo de análisis de sentimientos alcanzó una precisión del 77% en el conjunto de datos de prueba, lo que indica que es capaz de clasificar correctamente la mayoría de las opiniones en categorías positivas, negativas o neutrales.

3. Análisis crítico

El modelo alcanzó un porcentaje de precisión considerable, demostrando su capacidad para clasificar los sentimientos expresados en el texto. Esto sugiere que el modelo es efectivo en la tarea de análisis de sentimientos y puede ser útil en la identificación de opiniones positivas, negativas o neutras en textos en inglés.

Limitaciones identificadas:

Una de las limitaciones principales del modelo es su incapacidad para trabajar eficazmente con textos en idiomas distintos al inglés. Esto se debe a que el modelo fue entrenado exclusivamente con datos en inglés y, por lo tanto, puede no capturar adecuadamente las características lingüísticas y expresiones comunes en otros idiomas. Como resultado, los textos en español u otros idiomas pueden ser clasificados erróneamente como neutros, lo que limita la aplicabilidad del modelo en entornos multilingües.

Posibles mejoras:

Una mejora significativa sería la adaptación del modelo para trabajar con múltiples idiomas, lo que aumentaría su utilidad y aplicabilidad en entornos lingüísticamente diversos. Esto podría lograrse mediante la inclusión de datos en varios idiomas durante el entrenamiento del modelo y la exploración de técnicas de procesamiento de lenguaje natural que permitan la detección y clasificación efectiva de sentimientos en diferentes idiomas.

Enlace al código:

<https://github.com/LiamJimenez/EmocionesTexto.git>