

Trabajo del Tema 8

Trabajo del Tema 8

Descripción de la tarea: El csv "sentiment_dataset" incluye reviews que han dejado usuarios en la Google Play Store en referencia a 23 apps, incluyendo una valoración categórica sobre el análisis de sentimiento.

Los datos son los siguientes:

- package_name: Nombre de la app
- review: Texto del usuario con la review
- polarity: Sentimiento de polaridad de la review:
 - 1: Review con un sentimiento muy positivo o muy negativo
 - 0: Review con un sentimiento neutro

Se pide:

1. **Limpieza de instancias (3.5p):** Limpieza de texto de las instancias. Se valorará el uso de las técnicas empleadas y las explicaciones en el notebook
2. **Creación de modelo clasificador de Naive Bayes (3.5p):** Se valorará el uso de técnicas empleadas como el split entre entrenamiento y validación y selección de métrica de optimización
3. **Optimización del modelo (2p):** En base a los resultados obtenidos, ¿Somos capaces de mejorar nuestro modelo? Se valorará el uso de técnicas como Laplace smoothing, uso de la variable categórica package_name
4. **Conclusiones y vías abiertas (1p):** ¿Qué resultados obtenemos en el modelo?, ¿Predice correctamente?, ¿En todas las apps?, ¿Qué datos podrían ayudarnos a mejorar la predicción?

>> este ejercicio no tiene una solución única<< Pueden haber diversas opciones válidas, pero se espera que se haga una interpretación de los datos obtenidos en todos los puntos y que se construya un modelo de Naive Bayes acorde a lo analizado (Si la clasificación es válida, por qué lo es y si no llega a serlo, por qué no lo es y cómo podrían mejorarse los resultados)

Entrega:

- fichero Jupyter Notebook con MD004NombreApellidosAC08.ipynb
- fichero .pdf con MD004NombreApellidosAC08.pdf

Plazo: una semana (miércoles 07/02/2024 19:00).