



ANALISIS DE UNFPA - GRUPO 18

Santiago Osorio
Juan Andrés Eslava
Alejandro Segura





OBJETIVO

El objetivo de este proyecto es desarrollar y comparar modelos predictivos para clasificar textos en las categorías de los Objetivos de Desarrollo Sostenible (ODS) 3, 4 y 5, utilizando técnicas de aprendizaje automático. A través de esta clasificación, se busca facilitar la evaluación del cumplimiento de los ODS en Colombia, proporcionando información clave para mejorar la toma de decisiones y optimizar los recursos destinados a políticas de salud, educación y equidad de género.

ENFOQUE ANALITICO



El proyecto utiliza un enfoque predictivo basado en aprendizaje supervisado para clasificar textos en las categorías de los ODS (3, 4, y 5). Se implementaron tres modelos: Regresión Logística, Naive Bayes, y Árbol de Decisión, comparando su rendimiento mediante métricas de clasificación (precisión, recall, F1-score).

- Preprocesamiento de los datos
- División de los datos
- Entrenamiento de modelos
- Comparación de resultados
- Escoger mejor modelo
- Implementarlo y ver resultados



Regresión Logistica

Modela la relación entre variables explicativas y una variable binaria de resultado, permitiendo predecir la probabilidad de ocurrencia de un evento

f1-score: 0.76

MODELOS NAIVE BAYES

Utiliza la probabilidad condicional y el teorema de Bayes para clasificar datos, asumiendo independencia entre las características para predecir la probabilidad de pertenencia a una clase.

f1-score: 0.57

ARBOL DE DECISIÓN

Divide los datos en función de características mediante una serie de preguntas binarias, creando una estructura de árbol que clasifica los datos en diferentes categorías.

f1-score: 0.68

RESULTADOS

ODS 3 (SALUD Y BIENESTAR)

Métricas

- Precisión: 0.73
- Recall: 0.73
- F1-Score: 0.73

PALABRAS CLAVE:

- | | |
|-------------------|-------------------|
| Positivas: | Negativas: |
| • Maternal | • ciclista |
| • Sensibilizado | • divergent |
| • Dignidad | |

ODS 4 (EDUCACIÓN DE CALIDAD)

Métricas

- Precisión: 0.74
- Recall: 0.79
- F1-Score: 0.76

PALABRAS CLAVE:

- | | |
|-------------------|-------------------|
| Positivas: | Negativas: |
| • Importante | • Oncológica |
| • Adquirido | • Sacrificar |
| • Tolerada | |

ODS 5 (IGUALDAD DE GÉNERO)

Métricas

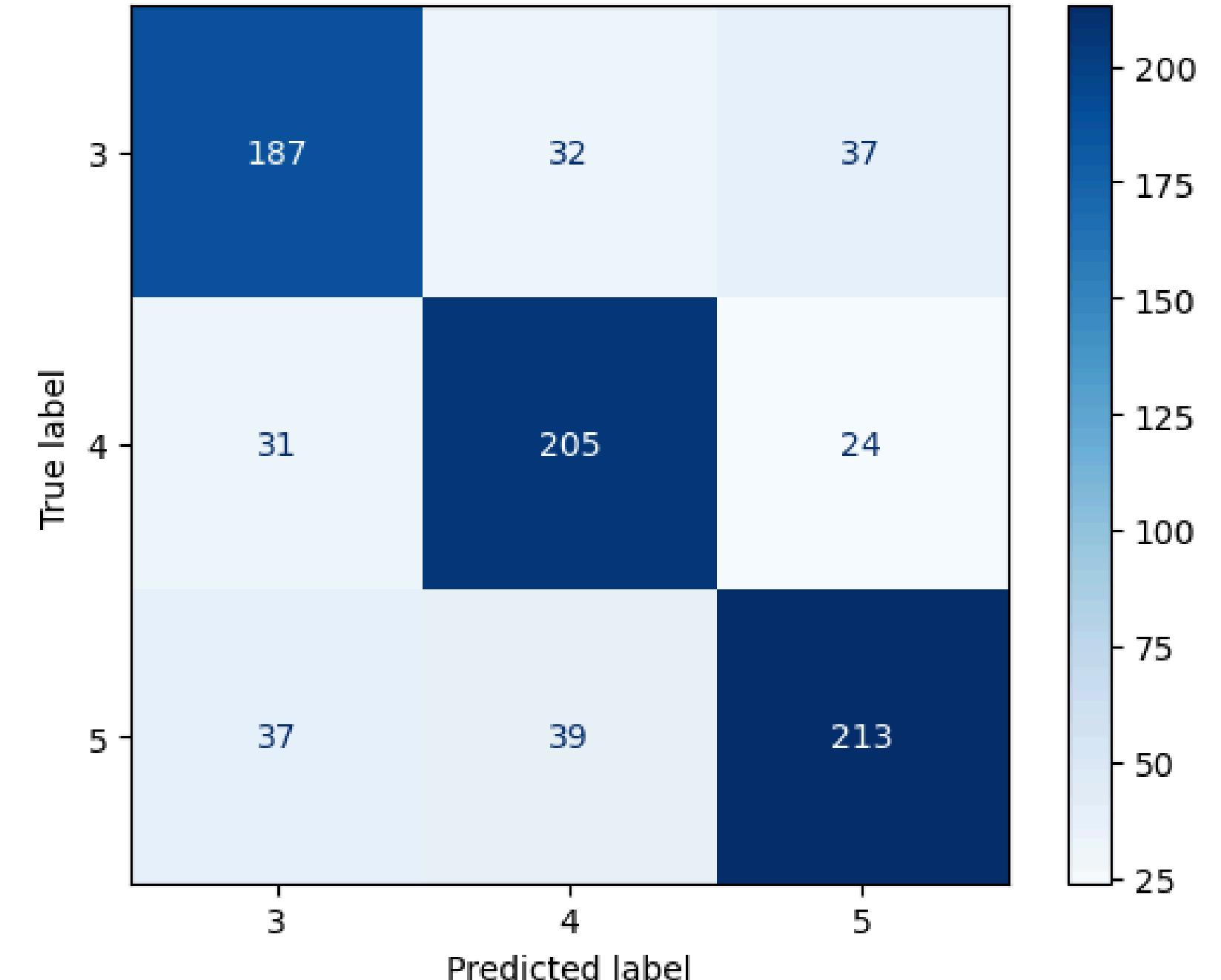
- Precisión: 0.78
- Recall: 0.74
- F1-Score: 0.76

PALABRAS CLAVE:

- | | |
|-------------------|-------------------|
| Positivas: | Negativas: |
| • Difundido | • Afectar |
| • Antibacteriano | • Recibir |
| • Lealtad | |

Matriz de Confusión

- Muestra un buen rendimiento en general para las 3 categorías, siendo 5 donde resalta más.
- En general hace un buen trabajo para predecir las categorías, evidenciandolo en su color azul oscuro.





CONCLUSIÓN

En base a los resultados, se recomienda el uso del algoritmo de regresión logística para la analítica de textos el Fondo de Poblaciones de las Naciones Unidas.

Además, la habilidad de clasificar de forma automática y precisa los comentarios facilita la identificación de fortalezas y áreas de mejora en la organización.

