



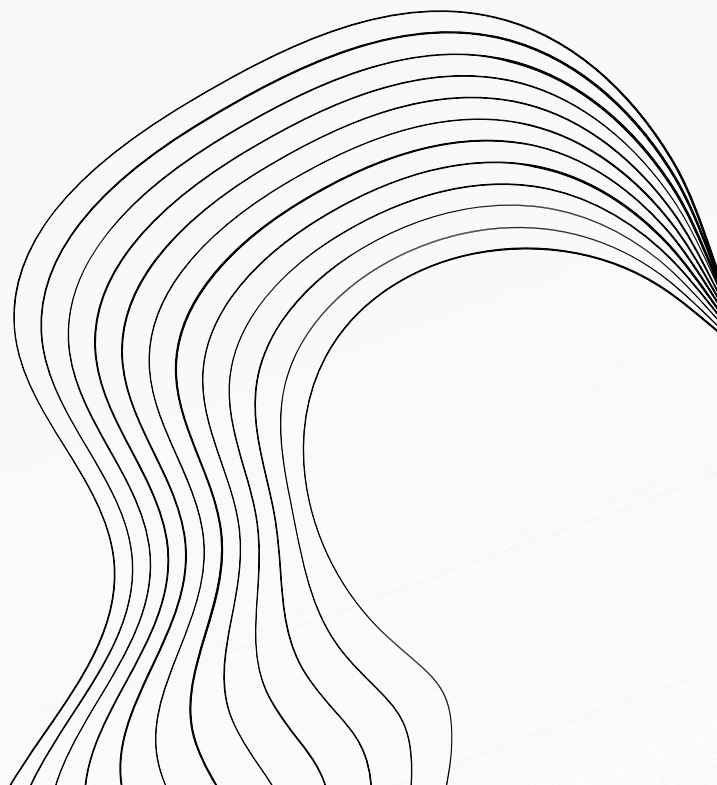
APRENDIZAJE AUTOMÁTICO PARA LA GESTIÓN DE DATOS MASIVOS

BY DR. PAULO LOPEZ MEYER

23/07/24

TAREA 09

**ISAAC
MENCHACA**



ANÁLISIS DE TEXTO

Este enfoque es útil en aplicaciones como la clasificación de documentos, la detección de spam y la categorización de contenido en plataformas de medios sociales. El resultado que da esta práctica con accuracy de 0.39493533208805237. Esto significa que el modelo predijo correctamente la categoría de aproximadamente el 39.49% de los textos en el conjunto de prueba. Tal vez utilizando un calificador diferente como SVM se obtendrán mejor accuracy, pero tardaría más en entrenarse, por lo que se optó por árboles.

✓ 2m 49.5s

Accuracy: 0.39493533208805237

Se hizo una prueba final introduciendo un nuevo texto que es el siguiente: "New breakthrough in cancer research". El calificador lo categorizo como 'e' que significa entretenimiento. De acuerdo al texto, debería caer más en el catálogo de ciencia y tecnología, por lo que se buscara mejorar el modelo de clasificación para obtener mejor resultado.

Predicted category: e