# Висновок

Під час виконання даної лабораторної роботи я ознайомитись з класифікацією документів за допомогою моделей машинного навчання.

Для початку я обробив набір даних: перетворив категоріальні змінні до числових, нормалізував тексти документів, використав TD-IDF модель для представлення тексту у векторному вигляді.

Після цього використав більше алгоритмів ніж було прописано у завданні, як-от: KNN, Logistic Regression, Random Forest, SVM, Naive Bayes.

Під час тренування моделей використовував GridSearch для підбору найкращих гіперпараметрів.

У результаті отримав, що на як на тестових, так і тренувальних даних SVM показала себе найкраще, однак до неї близько підібралися Logistic Regression та Naive Bayes. KNN гірше відпрацював на обох наборах ніж перші три, а Random Forest — найгірше на тестовому.