Зображення, що містить текст

Автоматично згенерований опис

Міністерство освіти і науки України

Національний технічний університет України

“Київський політехнічний інститут імені Ігоря Сікорського”

Факультет інформатики та обчислювальної техніки

Кафедра інформатики та програмної інженерії

**Лабораторна робота №4**

Аналіз текстів з використанням мови Python

**Тема:** Класифікація текстових даних

**Варіант:** 1

Виконав Перевірив:

студент групи ІП-11: Тимофєєва Ю. С

Панченко С. В.

Київ 2023

Table of Contents

[Мета лабораторної роботи 6](#__RefHeading___Toc3319_204808571)

[Завдання 7](#__RefHeading___Toc5937_3437555080)

# Мета лабораторної роботи

Ознайомитись з класифікацією документів за допомогою моделей машинного навчання.

# Завдання

Створити програму, яка зчитує заданий набір даних, виконує попередню обробку та класифікацію документів відповідно до варіанту. Якщо недостатньо ресурсів для роботи з повним набором даних, можна виділити частину, але таким чином, щоб були присутні усі класи.

Варіант 1.

Файл bbc-news-data.csv. В якості текстової моделі використати TD-IDF. Виконати класифікацію за допомогою алгоритмів наївний байєсів класифікатор та опорні вектори, порівняти їх точність. Спробувати покращити модель наївний байєсів класифікатор за допомогою GridSearchCV.