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 №3**

Аналіз текстів з використанням мови Python

**Тема:** Моделі 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 основними текстовими моделями та їх створення за допомогою scikit-learn та genism.

# Завдання

Варіант 1.

Зчитати файл doc1. Вважати кожен рядок окремим документом

корпусу. Виконати попередню обробку корпусу.

1. Представити корпус як модель «Сумка слів». Вивести вектор для слова «film».
2. Представити корпус як модель TD-IDF. Спробувати кластеризувати документи за допомогою ієрархічної агломераційної кластеризації.
3. Представити корпус як модель Word2Vec. Знайти подібні слова до слів shrimp, economy.