Московский государственный технический университет им. Н.Э. Баумана

Кафедра «Системы обработки информации и управления»

****

Лабораторная работа №9

по дисциплине

«Методы машинного обучения»

на тему

## **«Классификация текста»**

Выполнил:

студент Люй Чжэ

группы ИУ5И-23М

Москва — 2024 г.

**1. Цель лабораторной работы**

изучение методов классификации текстов.

**2. Задание**

Для произвольного набора данных, предназначенного для классификации текстов, решите задачу классификации текста двумя способами:

1.Способ 1. На основе CountVectorizer или TfidfVectorizer.

2.Способ 2. На основе моделей word2vec или Glove или fastText.

3.Сравните качество полученных моделей.

Для поиска наборов данных в поисковой системе можно использовать ключевые слова "datasets for text classification".

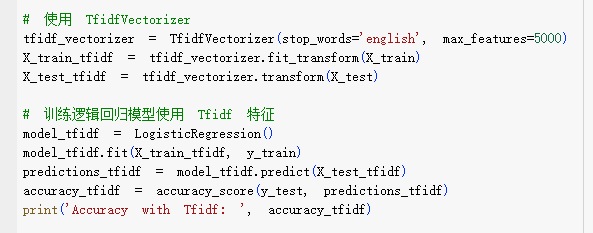
**3. Ход выполнения работы**

Я выбрал модель TfidfVectorizer и модель word2vec.

**Шаг 1: Загрузка данных**



**Шаг 2: Использование TfidfVectorizer**



**Шаг 3: Использование Word2Vec**



**Шаг 4: Сравнение результатов**

