Лабораторная работа 1 Word embeddings

Сравнить эффективность использования предобученных FastText и Word2Vec признаков с обученными самостоятельно. В работе необходимо использовать данные из корпуса Toxic Tweets Dataset

(https://www.kaggle.com/datasets/ashwiniyer176/toxic-tweets dataset)

Для решения задачи необходимо:

1: FastText и Word2Vec -признаки обучить на датасете Toxic Tweets Dataset.

2: Подобрать несколько вариаций использования различных параметров при обучении FastText и Word2Vec -признаков

3: Подгрузить предобученные FastText и Word2Vec -признаки. (https://fasttext.cc/docs/en/english-vectors.html)

4: Сравнить предобученные FastText и Word2Vec - признаки с лучшей вариацией самостоятельно обученных признаков с помощью самой эффективной модели классификатора из ваших работ предыдущего семестра.