# **Анализ текстовых данных средствами BERT и стека LSTM**

**Цель** работы – получить навыки обучения и использования архитектур BERT, стека LSTM и двунаправленной LSTM для задач классификации текстовых данных.

**Задачи** – Обучить нейросетевые модели BERT, стека LSTM, двунаправленную LSTM, предсказывающие авторство текстов по фрагменту текстов.

Для разработки можно использовать Tensorflow или PyTorch.

Выбрать 4-x поэтов из списка: Цветаева М.И., Некрасов Н.А., Лермонтов М.Ю., Тютчев Ф.И., Фет А.А., Бунин И.А., Есенин С.А., Маяковский В.В., Бродский И.А., Ахматова А.А., Блок А.А., Заболоцкий Н.А. (Можно взять любых других писателей). Для каждого выбранного поэта взять не менее 100 фрагментов из произведений, каждый фрагмент содержит не менее 100 символов.

Получить токенизатор и обученную модель BERT c HuggingFace.

Выполнить загрузку и предварительную обработку данных из текстов. Разделить каждую выборку на обучающую, тестовую и валидационную.

Добавить к архитектуре BERT полносвязный (или два полносвязных) слой(я) и обучить полученную модель в следующих вариантах:

1) Заморозить веса в BERT и настроить добавленный(е) полносвязный(ые) слой(и). (Это – модель 1а)

2) Выполнить Fine-tuning полученной архитектуры. (Это – модель 1b)

Обучить вторую модель - стек LSTM (две последовательные ячейки LSTM, добавить Dropout) для классификации текстовых данных.

Обучить третью модель – двунаправленную ячейку LSTM для классификации текстовых данных.

Привести кривые изменения функции потерь (Loss) и верности (Accuracy) во время обучения на обучающем и валидационном наборах для всех архитектур.

Привести результаты обучения моделей и сделать выводы.

**Описание отчета.**

1. Описание наборов данных.

2. Параметры архитектур и обучения нейронных сетей, использованные для обучения.

3. Графики обучения для архитектур нейронных сетей с лучшими характеристиками эффективности

4. Оценки моделей на тестовых выборках в виде таблиц/ диаграмм, отображающих метрики качества.

5. Программный код с комментариями.

6. Выводы