Отчет по практике Мурзовой Софьи ПМИ227

Задача: Обучить модель задачи классификации для определения, кому из персонажей Саус Парка больше подходит фраза. Определить зависимость модели от гипер параметров. Проанализировать, в каких случаях возникают ошибки.

Я использовала уже написанные Nlp модели из библиотеки fasttext (Стандартная, с оптимизированными гипер параметрами, измененным количеством эпох, с увеличенной скоростью обучения алгоритма, с оценкой определенного лейбла)

Сначала я загрузила датасет цитат персонажей со всех сезонов, затем, чтобы нормализовать текст, убрала заглавные буквы, знаки препинания и перенос строки. После этого я разделила все данные на обучающие и тренировочные и записала их в разные файлы. Дальше я попробовала обучить разные 5 моделей с измененными гипер параметрами на этих данных и выбрала модель, которая давала наибольшую точность(0.33) и сохранила ее, как самую оптимизированную

Результаты: Я реализовала решение задачи классификации фраз по персонажам саус парка с точностью 0.33. Наименее точная модель давала результат с точностью 0.25, поэтому, можно сделать выводы, что модель довольно сильно зависит от гипер параметров

Довольно сложно оценить ошибки…Это все же диалоги персонажей из Саус Парка, и понятие “Какой персонаж мог сказать эту фразу” довольно относительно, поэтому оценивать предсказания из произвольных данных будет некорректным. Но мне предсказания произвольных данных показались довольно реалистичными.

В тестовых данных были допущены ошибки, особенно часто это встречалось при классификации фраз редко встречающихся персонажей. Проблема этой ошибки связана с недостаточным количеством данных для обучения. Ведь для классификации высказываний главных героев было куда больше данных, чем для классификации второстепенных персонажей.

Выводы: Точность модели очень сильно зависит от гипер параметров. Наилучшая точность получилась у модели с увеличенной скоростью обучения алгоритма. Есть проблема решения классификации фраз персонажей- нехватка данных, для обучения на второстепенных героях, но довольно точно получилось реализовать классификацию по главным персонажам. По характеру высказывания модель определяет, кому она скорее относится.