

Яковлев Михаил

Telegram: @Michaelya
Phone: +7 953 448-77-48
Email: misha24-10@yandex.ru
Github: Misha24-10
Kaggle: mishayakovlev

ОБРАЗОВАНИЕ

НИУ «Высшая школа экономики», Москва 2019 - 2023
Бакалавриат. Информатика и вычислительная техника

НИУ «Высшая школа экономики», Москва 2023 - наст. время
Магистратура. Прикладные модели искусственного интеллекта

ОПЫТ РАБОТЫ

НИУ «Высшая школа экономики», Москва Январь 2023 - Май 2023
Учебный ассистент по дисциплине: «Математика для анализа данных»

Инженерно-математическая школа VK и НИУ ВШЭ
Октябрь 2023 - наст. время
Сбор датасета для генерации речи, собрано более 10 тыс. часов. Обучение моделей TTS на больших данных.

ПРОФ. НАВЫКИ

Python (продвинутый), C++ (Основы), Pandas, Numpy, Matplotlib, Seabon, Catboost, Xgboost, NLP, CV, RecSys, Pytorch, Keras, Transformers, Статистика, Линейная алгебра, SQL, Linux, Git

ЯЗЫКОВЫЕ НАВЫКИ

Английский - B1

ПРОЕКТНЫЙ ОПЫТ

Соревнование Kaggle
Bengali.AI Speech Recognition
Для решения задачи OOD (out-of-distribution) распознавания речи для бенгальского языка была обучена модель Wav2Vec2. Также была попытка дообучения модели Whisper, но данная модель показала себя хуже по сравнению с Wav2Vec2. WER для итоговой модели составило 52%, данное решение заняло 91/745 место (бронза).

SemEval Workshop

Multilingual Complex Named Entity Recognition

Для решения задачи мультязычного распознавания именованных сущностей на 12 языках было проведено дообучение нескольких мультязычных моделей, включая mBERT, RoBERTa и RemBERT. Благодаря ансамблированию дообученных языковых моделей удалось достичь качества $F_1 = 0.734$.

Учебный проект НИУ ВШЭ

Использование алгоритмов компьютерного зрения для аутентификации лица пользователя с камеры ноутбука.

В рамках данного проекта была реализована модель для распознавания лиц, для этого использовалась архитектура MTCNN+ResNet50+ArcFace. Модель была протестирована на специально собранном для тестирования датасете и продемонстрировала высокую точность распознавания с $FPR < 1\%$ и $FNR < 2\%$.

Обучение переводчика на основе модели T5

В рамках данного проекта была обучена модель машинного перевода на основе mT5. Для обучения модели использовался параллельный корпус текстов на двух языках турецкий и английский. Обученная модель смогла достичь качества BLEU=0.22.