

Бородин Матвей Вячеславович

+7 916 715 9563 | salam121asd@gmail.com | t.me/Djatva | github.com/matveylogeer

Образование

НИУ ВШЭ

Бакалавриат, Факультет компьютерных наук, Программная инженерия

Москва

2022 — 2026

Олимпиадный опыт

Победитель Всесибирской открытой олимпиады школьников по математике 2020 — 2021

Призер олимпиады «Шаг в будущее» по математике 2020 — 2021

Курс «Олимпиадные задачи» Московской Школы Программистов 2020 — 2022

Олимпиадная школа МФТИ по математике 2019 — 2020

Релевантный опыт

А/В Week Школы Анализа Данных | Яндекс Июль 2025

Тренировки. ML-забег | Яндекс Сентябрь 2025

НИС Технологии Компьютерного Зрения | НИУ ВШЭ, ФКН Сентябрь 2025 — Наст. время

Курс Математические Методы Анализа Данных | НИУ ВШЭ, ФКН 2024 — 2025

Научно-практический интенсив по мобильной разработке | Сириус, Яндекс Март 2025

- Поработал в команде из 5 человек над приложением для фестиваля YoungCon (описание ниже).

Реализованные проекты

Классификация кризисно-релевантных твитов | LSTM+word2vec, BERT, XLNet, Fine-tuning

- Цель проекта — классификация твитов из датасета **CrisisLexT26**, определять, содержит ли сообщение информацию, полезную при кризисах (пожары, землетрясения и тд.) или нет.
- Предобработка и данные — выполнена нормализация текстов, перевод, удаление дубликатов и сплит на train / validation / test с учётом стратификации.
- Исследованы три подхода: **LSTM + Word2Vec** (из базовых эмбеддингов) и дообучение трансформеров: **BERT** (bert-base-uncased), **XLNet** (xlnet-base) с **fine-tuning** и подбором гиперпараметров.

Классификация отзывов по категориям | XLM-RoBERTa, LoRA, PEFT, Zero-shot Weak-labeling

- Проект решает задачу многоклассовой классификации отзывов товаров на маркетплейсах — присвоить каждому отзыву одну из фиксированных категорий с высокой метрикой **Weighted F1**.
- Сначала используется **zero-shot** модель **XLM-RoBERTa-large-XNLI** для **weak-label** разметки, затем отбор надёжных примеров по уверенности, балансировка классов и аугментация.
- К основной модели классификации **XLM-RoBERTa-base** применены **PEFT / LoRA** (адаптеры к основным слоям q/k/v/dense) техники дообучения.

Kaggle NYC Taxi Trip Duration | EDA, Feature Engineering, sklearn, Linear Regression

- Задача предсказать длительность поездки на NYC такси на основе данных о заказах (координаты, время, и др.), минимизировав метрику **RMSLE**.
- Проведён **EDA** и **Feature Engineering**: лог-трансформация таргета, чистка выбросов и аномалий, скейлинг признаков; временные и геометрические признаки (расстояние по **Haversine**, логарифм расстояния), пространственная дискретизация через сетку **GridBinner**, «горячие зоны» через **KMeans**, **one-hot** кодировки.
- Для предсказаний использовались линейные **Ridge** и **Lasso** модели в рамках **sklearn**-пайплайнов.

iOS-Приложение YoungCon | Swift, SwiftUI, UIKit, MVVM+Coordinator, TestFlight, SwiftLint

- Разработано приложение YoungCon для одноименного фестиваля на фреймворке **SwiftUI** с применением **UIKit** при помощи **UIViewRepresentable**, архитектурой **MVVM+Coordinator**, бэкендом на **Golang** и хранением данных в **Google Таблицах**. Больше разраб проектов в мое гите.

Технические навыки

Языки: Python, C++, Swift, Kotlin, Java

Технологии: Numpy, Pandas, PyTorch, SciPy, sklearn, transformers, keras-preprocessing, XGBoost, CatBoost, MLflow, FastAPI, LoRA, PEFT, NLP, CV

Софт: Git, Docker, Postman, Collab, Figma, TestFlight, LaTeX