

Илья Грибов

ML-engineer, Data Science

✉ ai.data.explorer@gmail.com | ✉ mathandmlexplorer@yandex.com | 📞 +7 936 133 68 39
🌐 UniverselsMyCreation | 🐙 @math_ml_explorer

Образование

МГУ им. М.В.Ломоносова

Дата выпуска: Июнь 2024

Бакалавр: Прикладная математика и информатика

Средний балл: 4.86

Пройденные курсы: **Математические:** Математический анализ, Линейная алгебра, Теория вероятностей и Математическая статистика, Дискретная математика, Случайные процессы, Численные методы, Дифференциальные уравнения, Методы оптимизации; **ML&DS:** Машинное обучение, Глубокое обучение, Методы обработки и распознавания изображения (ШАД), Математические методы распознавания образов, Анализ неструктурированных данных (тексты, картинки); **CS:** Архитектура ЭВМ и Операционные системы, ООП(C++, Python), Алгоритмы и структуры данных, Введение в сети ЭВМ, Базы данных, Распределенные системы и параллельное программирование

Дополнительное образование: **VK:** Введение в рекомендательные системы, Информационный поиск, **DLS:** Компьютерное зрение, Обработка текстов и звука, **YSDA:** Введение в LLM

Опыт работы

Тинькофф Банк

Октябрь 2023 - По настоящее время

ML-engineer(NLP, classical ML)

Python, Git, SQL, PyTorch, ONNX, Sklearn, SpaCy, Scipy, Seaborn e.t.c.

- Анализ и обработка исходных данных; Разработка и внедрение ML и DL моделей для прикладных задач по обработке текста.
- Проведение экспериментов по подбору оптимальной архитектуре модели, общего пайплайна работы моделей, ускорению работы модели (в т.ч. с использованием ONNX), интерпретации модели (в т.ч. с использованием LIME, ALTI, SHAP, Gradients).

Персональные проекты

Decision Tree

Реализация алгоритма решающего дерева на языке C++

C++

- Был реализован алгоритм обучения и inference решающего дерева

RecSys approaches

Реализация и исследование основных подходов RecSys

Python, Numpy, Pandas, Sklearn, PyTorch, Git

- Коллаборативная фильтрация
- ALS и deep learning
- user2user, item2item, user2item
- Матричная факторизация

Adversarial attacks & Interpretation (Research work)

Исследование методов порождения состязательных атак

Python, Jax, Transformers, PyTorch, Git

Captum, Lime, Shap e.t.c.

- Изучение статей по методам генерации состязательных атак и по интерпретации предсказаний моделей
- Реализация pipeline для генерации состязательных атак от начала и до конца (с учетом извлечения интерпретации)
- Подготовленный отчет об результатах экспериментов и исследований

Conformer architecture for Audio Denoising

self-made Conformer model for audio denoising

Python, PyTorch, TorchAudio, Glob, Librosa, Git

- Подготовка датасета для зашумления исходных чистых аудио шумом, реализация основных метрик, предобработка звука
- Реализация Conformer model для данной задачи, в т. ч. включающего Multi-Headed Self-Attention с Relative positional encoding и RelativeShift
- Обучение модели, замер итогового качества

Competitions and Hackathons

VK TLS-embedding

May 2023

Соревнование по построению качественного TLS-embedding

- Необходимо было по данным в датасете TLS токенам и разметке spam/not-spam обучить TLS-embedding, чтобы фильтровать ботов
- В процессе решения задачи использовались основные подходы для токенизации текста (fasttext, tokens, BPETokenizer, e.t.c.) и обучения векторных представлений (word2vec, bert, rnn, cnn, e.t.c.)

Kaggle

2023

Участие в соревнованиях на Kaggle

- Участие в соревновании на Kaggle по предсказанию цены поездки такси. При решении задачи проводилось EDA (exploratory data analysis), и грамотный feature engineering, итоговой моделью использовался градиентный бустинг
- Участие в соревновании на Kaggle по предсказанию активности пользователей Steam, в котором для достижения наилучшего результата проводился feature engineering и feature selection, а также использовался Stacking алгоритмов машинного обучения

Hard Skills

Programming languages:

C++, SQL, Bash, Python

ML & DS frameworks:

DS: numpy, pandas, seaborn, matplotlib, scipy e.t.c.

ML: scikit-learn, imblearn, catboost, xgboost, optuna, hyperopt e.t.c.

DL: pytorch, onnx, tensorflow (beginner), jax, transformers e.t.c.

Computer Vision: opencv, torchvision, albumenations, PIL, scikit-image, GAN, CNN e.t.c.

NLP: torchvocab, nltk, re, natasha, spacy, razdel, RNN, GPT, BERT e.t.c.

Audio Procesing: librosa, torchaudio e.t.c.

Technologies & Tools:

FastAPI (beginner), Git, Linux, Docker (beginner), WandB, Jira, Confluence

Languages:

Russian (native), English(B2-C1), Germany(A1)