

Илья Грибов

ML-enginner/researcher, DataScience

✉ ilyagr.ml.rnd@gmail.com | ✉ ilyagr.ml.rnd@yandex.ru | 📞 +7 (936) 133 68 39

🔗 Math-AI-explorer | 🐙 @math_ml_explorer

Education

МГУ им. М.В.Ломоносова

Сентябрь 2020 - Июнь 2024

Бакалавр: Прикладная математика и информатика

Средний балл: 4.77

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;

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, Введение в LLM; **DLS:** Компьютерное зрение, Обработка текстов и звука

МГУ им. М.В.Ломоносова

Сентябрь 2024 - Июнь 2026

Магистр: Прикладная математика и информатика

Средний балл: 4.9

Пройденные курсы:

Компьютерное зрение, Современные методы обработки изображений, Глубокое обучение, Прямой и Обратный рендеринг, 3D Компьютерное зрение

Experience

Тинькофф Банк

Октябрь 2023 - Март 2024

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).

Сбер

Март 2024 - Октябрь 2024

ML-researcher (RecSys)

Python, Git, Bash, PyTorch, Polars, PySpark, CatBoost, Sklearn, Jax e.t.c

- Разработка Foundation Model: построение архитектур моделей, выдвижение гипотез и их проверка, чтение и реализация статей, обработка данных для обучения модели;
- Участие в соревновании RecSys Challenge 2024 по рекомендации статей в электронном журнале: генерация фичей, выдвижение и проверка гипотез, обучение различных моделей на различных задачах.

Hard Skills

Languages:

Python, C++/C, Go, SQL, Bash

Technologies & Tools:

Poetry, FastAPI, SQLAlchemy, Git, Linux, Docker, WandB, Aim, Jira, Confluence

ML & DS frameworks:

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

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

DL: pytorch, onnx, tensorflow, torch-pruning, jax, flax, transformers e.t.c.

CV: opencv, torchvision, albumenations, PIL, scikit-image, GAN, CNN, Duffusion e.t.c.

3D CV: open3d, COLMAP, SLAM, NeRF, 3D Gaussian Splatting e.t.c.

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

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

RL: pytorch, gymnasium, Q-Learning, policy-gradient, A2C

Languages:

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

Projects

Decision Tree

Jul. 2022

Реализация алгоритма решающего дерева на языке C++

C++, Git

- Был реализован алгоритм обучения и inference решающего дерева

RecSys approaches

Aug. 2023

Реализация и исследование основных подходов RecSys

Python, Numpy, Pandas, Sklearn, PyTorch, Git

- Коллаборативная фильтрация
- Deep Learning
- user2user, item2item
- ALS и Матричная факторизация

Adversarial attacks & Interpretation (Research work)

Mar. 2024

Исследование методов порождения состязательных атак

Python, Transformers, PyTorch, Git, Lime e.t.c.

- Изучение статей по методам генерации состязательных атак и по интерпретации предсказаний моделей
- Реализация pipeline для генерации состязательных атак с использованием интерпретации
- Подготовленный отчет об результатах экспериментов и исследований

Conformer architecture for Audio Denoising

Oct. 2023

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