



Марченко Андрей
Junior Data Scientist
21 год

+7-967-123-25-25
marcheanin1576@gmail.com
GitHub Profile

Профиль

• Младший Data Scientist / ML-инженер с опытом разработки production-микросервисов на Python (FastAPI) и интеграции LLM (Yandex Cloud) для задач кредитного скоринга, генерации и оценки тестов и AI-интервью. Опыт построения end-to-end ML-пайплайнов (предобработка, обучение, деплой, оценка качества) и работы с семантическим поиском (FAISS, эмбединги) и RAG-подходами.

Образование

- **МГТУ им. Н. Э. Баумана** 2021–2025
Бакалавриат, ИУ9 Теоретическая информатика и компьютерные технологии (ПМИ) Москва
ВКР: Разработка алгоритма автоматического формирования модели кредитного скоринга
- **МФТИ** 2025–2027
Магистратура, ПМИ, ИВТ, Кафедра банковских информационных технологий Москва
- **VK Education – Отделение при МГТУ им. Н.Э. Баумана** 2023–2024
Программа: Машинное обучение. Основные курсы: базовый ML, Deep Learning, DWH


Проекты

- **AutoML для кредитного скоринга** Дипломная работа, Весна 2025 [Source Code](#)
 - Система автоматического формирования полного пайплайна для задач бинарной и многоклассовой классификации в кредитном скоринге.
 - Оптимизация на основе генетического алгоритма, кодирующего 22-гены хромосомы пайплайна.
 - Включает до 5 вариативных этапов предобработки данных и 4 модели машинного обучения.
 - Реализован удобный интерфейс CLI и сериализация пайплайна для дальнейшего использования.
 - **Используемые технологии:** Python, Scikit-learn, Pandas, Numpy, Imbalanced-learn, Gensim, TensorFlow.
- **Анализ мемов на hateful содержание** Pet-проект, Декабрь 2023 [Source Code](#)
 - Сервис, анализирующий мемы (картинка+текст) на ненавистническое содержание.
 - Разработал объединение сверточной и текстовой моделей с помощью классификационной головы.
 - **Используемые технологии:** Python, Pandas, PyTorch, Lightning.
- **ClusterText** Pet-проект, Весна 2024 [Source Code](#)
 - Web-приложение для кластеризации текстов, выделения ключевых слов и расстановки тегов.
 - Разработал модель для кластеризации текстов, относящую текст к одному или нескольким из 15 кластеров.
 - Модель состоит из эмбеддера на основе SBert и KMeans для обучения на 60k статьях.
 - **Используемые технологии:** Python, Pandas, Clustergram, KMeans, Transformers, PyTorch.
- **Многокритериальная оптимизация в задаче кредитного скоринга** Курсовой проект, Декабрь 2024 [Source Code](#)
 - Построение многокритериальной модели для оптимизации разноплановых loss-функций.
 - Применение генетических алгоритмов для обработки гетерогенных данных и разработка собственного алгоритма оптимизации.
 - **Используемые технологии:** Python, Pandas, TensorFlow, PyTorch.
- **Codeix: генерация тестов и AI-интервью** Стартап-проект, Зима 2026 [Tests Service](#)
 - Набор микросервисов для обучающей платформы: генерация тестов по шаблону/теме/тегам и проведение технического AI-интервью.
 - Тесты генерируются по материалу, пройденному учеником, либо расширяют эталонные тесты; реализован LLM-as-a-judge для оценки качества сгенерированных тестов.
 - Интервью: вопросы формируются на основе RAG и/или заданного контекста, ответ записывается голосом, транскрибируется и анализируется через LLM (Yandex Cloud).
 - Реализована постобработка результатов STT (ASR error correction) и LLM-as-a-judge для оценки интервью по критериям.

- Микросервисы используют Yandex Cloud API для LLM-моделей, распознавания речи и объектное хранилище; реализованы на асинхронном FastAPI.
- **Используемые технологии:** Python, FastAPI, LLM, RAG, Yandex Cloud, API-вызовы, async

- **Agents-itry**

Pet-проект, Зима 2026

 [Source Code](#)

- Pet-проект для тренировки навыков разработки AI-агентов.
- На основе LangChain разрабатываю агентов с tools для поиска в интернете, локальных расчетов арифметики, изменения кода.
- **Используемые технологии:** Python, LangChain, LLM, API-вызовы, async

ОПЫТ

- **Компания ИНВЕРСИЯ**

Июль 2022 – Февраль 2023

Программист, Java, PL/SQL

Москва

- Работал в связке Java – SQL, решал задачи как во фронтенд, так и в бэкенд частях.

- **Газпромбанк**

Июнь 2024 – наст. время

Инженер разработки, Java

Москва

- Микросервисная разработка на языке Java.
- Разработка сложных структур баз данных PostgreSQL.

- **СберТех**

Июнь 2024 – наст. время

IT-инженер, ML

Москва

- Работал над задачами, связанными с оптимизацией хранения данных через дедупликацию путём семантического поиска.
- Разрабатывал полный пайплайн такого решения, который состоит из предобработки данных, выделения групп кандидатов через FAISS-поиск ближайших соседей, feature-engineering и применения моделей градиентного бустинга для классификации.
- Результаты были оценены через LLM-постобработку FP-пар, которая показала высокую эффективность разработанного метода.

ТЕХНИЧЕСКИЕ НАВЫКИ

Языки программирования: Python (продвинутый уровень, опыт в разработке production-решений), SQL (написание сложных запросов с использованием оконных функций и CTE), C++ (опыт спортивного программирования и разработки проектов), Java (опыт коммерческой разработки).

Библиотеки и фреймворки: PyTorch, Scikit-learn, Pandas, Numpy, Matplotlib, Seaborn, Transformers, Lightning, Imbalanced-learn, Gensim, TensorFlow, FastAPI, LangChain.

Инструменты: Git, Docker, VS Code, PyCharm

Навыки: Машинное обучение, глубокое обучение, NLP, компьютерное зрение, семантический поиск (FAISS, эмбединги), RAG, A/B тестирование, статистический анализ, алгоритмы и структуры данных.

Языки: Русский (родной), Английский (C1), Немецкий (A2)

СИЛЬНЫЕ СТОРОНЫ И ИНТЕРЕСЫ

Сильные стороны: усердие, исполнительность, ответственность, трудолюбие, способность к обучению и развитию, чувство юмора

Увлечения: настольный теннис, шахматы (I взрослый разряд), футбол, робототехника.