

# Апухтин Владислав Борисович

ML Engineer / Data Scientist (Intern/Junior)

+7-915-062-2993 | apukhtin00@inbox.ru | Москва | Tg: @kyl0q | github.com/V-L-A-P-P

## Summary

Студент направления «Прикладное машинное обучение». Разрабатываю высоконадёжные ML-решения: от подготовки данных и построения моделей до развёртывания сервисов (FastAPI + Docker). Реализовал проекты в предсказательной аналитике, NLP, RAG и рекомендательных системах. Участвую в олимпиадах, хакатонах, публикую статьи по ML. Нацелен на развитие и решение прикладных ML-задач в команде профессионалов.

## Опыт работы

### Algorithm Engineer (Project-based)

RWB — оптимизация вещания рекламных кампаний (в рамках KIM Project)

Сентябрь 2025 – н.в.

Москва

- Разрабатывал алгоритм оптимизации расписания рекламного вещания: учитывал бизнес-ограничения, овербукинг и обеспечивал равномерное распределение показов по слотам/времени.
- Результат:** Разработал алгоритм генерации расписания; уменьшил долю ручных правок на 50%.

## Проекты

Май 2025 – Июнь 2025

### Bankruptcy Prediction System — production-ready ML-сервис для оценки риска банкротства

Python | Pandas | Scikit-learn | CatBoost | Optuna | SHAP | FastAPI | PyTest | Docker

- Собрал и структурировал крупный финансовый датасет (≈700k компаний, 110 признаков) с выраженным дисбалансом классов; провёл глубокий **EDA** и **feature engineering** (125 доменных и производных признаков).
- Построил модели, отобрал лучшую и настроил гиперпараметры через Optuna с учётом сильного дисбаланса (1/250); финальная модель Catboost достигла **weighted accuracy = 0.87**.
- Окалибровал вероятности с помощью Isotonic Regression и внедрил интерпретацию модели через **SHAP**.
- Реализовал **полный ML-пайплайн** обучения и инференса, тестирование (**PyTest**) и деплой сервиса (**FastAPI**, **Docker**).

Август 2025 – Сентябрь 2025

### AURA — интеллектуальная RAG-система поиска и генерации ответов по финтех-корпусу

Python | FAISS (HNSW) | Sentence-Transformers | Cross-Encoder | Graph Ranking | LLM | NetworkX | FastAPI | Docker

- Реализовал **гибридный retrieval** (TF-IDF + FAISS + двухэтапный cross-encoder reranking на **transformers**), обеспечив прирост качества относительно базового RAG.
- Внедрил **LLM-based query rewriting** на базе instruction-tuned моделей для повышения релевантности поиска.
- Построил пайплайн подготовки данных с несколькими стратегиями чанкинга (базовый, семантический, рекурсивный) и предварительным препроцессингом документов.
- Разработал **граф связей между чанками** и интегрировал его в retrieval-этап для graph-augmented расширения кандидатов и reranking.
- Реализовал **end-to-end пайплайн RAG** с **LLM answer generation**: ответы по найденным документам с указанием источников; тестирование и деплой (**PyTest**, **FastAPI**, **Docker**).

## Movie Recommendation System – ML-модель персональных рекомендаций фильмов

Python | Pandas | NumPy | SciPy | Scikit-learn | Implicit ALS | Collaborative Filtering | Ranking Metrics

- Разработал персонализированную рекомендательную систему фильмов с использованием **User-Based CF**, **Item-Based CF** и **confidence-weighted implicit ALS**.
- Реализовал **cold-start handling** и **explainable рекомендации**: явное объяснение выбора фильмов через вклад похожих ранее просмотренных items (item-item similarity).
- Спроектировал **единый evaluation pipeline** и провёл сравнение моделей по **Recall@10**, **NDCG@10** и **MAP@10**

## Навыки

---

- **Языки и инструменты**: Python, SQL, Git, Bash, Linux
- **ML и анализ данных**: Pandas, NumPy, Scikit-learn, CatBoost, XGBoost, LightGBM, Optuna, SHAP
- **Deep Learning**: PyTorch, Transformers, RAG, HuggingFace
- **MLOps и сервисы**: FastAPI, Docker, PyTest
- **Математика**: линейная алгебра, матанализ, теория вероятностей, математическая статистика

## Образование

---

Финансовый университет при Правительстве РФ

2024 – 2028

Факультет информационных технологий и анализа больших данных

Москва

Направление/профиль: прикладная математика и информатика/прикладное машинное обучение

Курсы:

- |  |      |
|--|------|
| • Практический Machine Learning (Сертификат с отличием) – AI Education                 | 2025 |
| • Продвинутые методы машинного обучения (Сертификат с отличием) – AI Education         | 2025 |
| • Практический Deep Learning (Сертификат с отличием) – AI Education                    | 2025 |
| • Введение в Data Science и машинное обучение (Сертификат с отличием) – Karpov Courses | 2025 |

## Дополнительно

---

- Активно участвую в хакатонах и соревнованиях по анализу данных, помогаю в организации хакатонов
- Пишу статьи и регулярно изучаю новые подходы в ML/Deep Learning.
- Языки: Русский – Native, Английский – B2, Немецкий – A2
- Личные качества: ответственность, аналитическое мышление, стрессоустойчивость, коммуникабельность