

# Медведев Давид Валерьевич

## ML-Инженер

После инженерного проекта по автоматическому расчёту инсоляции помещений (степень освещенности помещений солнечным светом), основанного на методах линейной алгебры, я всерьёз увлекся математикой, что в конечном итоге привело меня в МФТИ и к сфере анализа данных и машинного обучения. Хочу развиваться Data Science и ML-инжиниринга, сосредотачиваясь на прикладных задачах. Также рассматриваю возможность поступления в аспирантуру и участия в исследовательских проектах.

---

## КОНТАКТЫ:

**Telegram:** @david\_medvedev

**GitHub:** <https://github.com/SalLangg>

**Интерактивное резюме:** <https://sallangg.github.io>

---

## ОБРАЗОВАНИЕ:

2024 — *по настоящее время*

**МФТИ, ФПМИ** — Магистратура

**Факультет:** Прикладная математика и информатика

**Направление:** Современная комбинаторика

2018 — 2022

**Тюменский индустриальный университет** — Бакалавриат

**Направление:** Расчеты строительных конструкций

---

## ОБУЧЕНИЕ И КУРСЫ:

- **Центр "Пуск" МФТИ:** Продвинутые методы машинного обучения.
- **Кафедра интеллектуальных систем:** Введение в машинное обучение (Константин Воронцов) и программирование на Python (Мурат Апишев)

---

## ПРОЕКТЫ:

### BuildBIMClassify — ML-система классификации BIM-объектов по сметным работам.

Разработал и внедрил пилотную MVP модель для автоматического сопоставления элементов BIM-модели (Revit) с позициями сметного справочника - классическая задача, выполняемая по большей мере вручную.

**Цель:** сократить до минимума ручное сопоставления элементов по работам. Когда сбор справочника будет завершен, модель будет полностью обучена и внедрена в боевые проекты, что позволит значительно сократить ручной труд и повысить точность сметных расчётов.

**Технологии:** Scikit-learn, LightGBM, CatBoost

#### Описание:

- Использовал CatBoost для multi-label классификации с фокусом на Precision (90%).
- Построил пайплайн сбора данных, их обработки и использования результатов модели.
- Подготовил пайплайн для дообучения модели при расширении справочника.

Сейчас работаю над повышением качества классификации к уровню 95-98%. Дальнейшая цель - настроить модель так, что бы нивелировать грамматические ошибки в текстовых переменных.

### **MorseNet — Декодер аудио файлов с кодом Морзе**

**Цель:** построить модель декодирования сигналов морзе, используя технологии, похожие на обработку естественного языка.

**Технологии:** PyTorch, FastAPI, MLflow, Docker

**Описание:**

- Построена нейросеть *CNN → LSTM* с *CTC Loss* для декодирования из аудиофайлов
- Для извлечения признаков использовались *Mel-спектрограммы* и аугментации (*time/freq masking*)
- Логирование метрик обучения происходит через *MLflow*
- Сохранение моделей на сервере
- Качество: **0.433 no Levenshtein distance** на Kaggle (15 место, лидер — 0.24)
- Реализован *FastAPI-сервер* с возможностью дообучения модели, независимо от инференса
- Решение упаковано в *Docker*

**GitHub:** [https://github.com/SalLangg/Morse-Decoder\\_V2](https://github.com/SalLangg/Morse-Decoder_V2)

### **Классификация изображений**

**Цель:** построить модель для классификации 42 персонажей по JPEG-изображениям

**Технологии:** PyTorch, torchvision, seaborn

**Описание:**

- Разработана *CNN-модель* с 3 сверточными блоками
- Использованы техники аугментации и расширения тестовой выборки: случайные повороты, изменение яркости/контраста, горизонтальное отражение
- **96.56% accuracy** на тренировочной выборке для самой базовой модели.

**GitHub:** [https://github.com/SalLangg/Personality\\_Prediction](https://github.com/SalLangg/Personality_Prediction)

### **Предсказание личности**

**Цель:** Цель: разработать модель предсказания личности человека на основе данных

**Технологии:** Pandas, Numpy, Matplotlib, Seaborn, StratifiedKFold, CatBoostClassifier, scikit-learn

**Описание:**

- Проанализирована степень важности пропусков в данных
- Созданы новые признаки для расширения выборки
- Протестированы различные модели - *CatBoost, XGBoost, RandomForest*, а также их стейкинз

**GitHub:** <https://github.com/SalLangg/Image-classification>

### **RAG-LMM помощник инженера (В разработке)**

**Цель:** разработать систему умного поиска по внутренней базе знаний компании с выводом найденной информации в качестве контекста LLM.

**Технологии:** модели компьютерного зрения, ORC, LLM, RAG

**Задачи, которые предстоит решить:**

- Решение проблем с неструктурированной документацией с помощью *ORC*
- Интеграция примечаний с изображениями в систему
- Проверка актуальности норм
- Фильтрация галлюцинаций LLM

---

## **ПРОФЕССИОНАЛЬНЫЙ ОПЫТ:**

**ВМ-Менеджер / ООО «Партнер.Проект»**

**02/2023 — по настоящее время**

*В рамках одной должности прошел путь от технической поддержки проектов до разработки ML-решений.*

**Data Scientist (2025 — настоящее время)**

- Разработал MVP-модель для сопоставления элементов BIM-модели (Revit) с позициями сметного справочника
- Использовал LightGBM для multi-label классификации, целевая метрика — Precision 90%
- Искусственно вводил пропуски в обучающую выборку для повышения устойчивости к человеческому фактору
- Построил пайплайн дообучения модели при расширении справочника
- Сейчас — работаю над улучшением качества до стабильных 95–98%. Затем заменю категориальные признаки на эмбединги (устойчивость к опечаткам) и разверну сервер для обработки данных.

**Цель проекта:** повышение точности расчетов путем уменьшения значимости ошибок при ручном заполнении параметров в информационной модели.

**Python-разработчик BIM (2024—2025)**

- Разработал сложные инструменты автоматизации, интегрированные в ПО для проектирования. Разработка только 2-х плагинов помогли сэкономить 550 тыс. руб. в год.
- Собираю и анализирую данные с разработанных плагинов и информационных моделей. Это позволяет найти и закрыть слабые стороны проектирования
- Собираю и выгружаю данные из BIM-моделей в PostgreSQL по запросам аналитиков

**BIM-Координатор (2023 — 2024)**

- Обеспечиваю полную техническую поддержку проектной группы по вопросам BIM
- Разрабатывал простые средства автоматизации проектирования (скрипты, шаблоны, макросы)
- Создавал обучающие материалы для сотрудников, ~40 человек (видеоуроки, база знаний)

---

**НАВЫКИ:**

**Языки программирования:** Python, C# (базово)

**ML:** PyTorch, sklearn

**Прочее:** SQL, Docker, FastAPI, MLflow,