



+7(915)140-35-97  
[malkinavladushka28@gmail.com](mailto:malkinavladushka28@gmail.com)  
<https://github.com/MalkinaVlada>

## ОПЫТ ЛИЧНЫХ ПРОЕКТОВ

### Выпускной проект на курсе Samsung Innovation Campus «Big Data», Июнь 2024

Построила конвейер данных для обработки датасета о клиентах банка в потоковом формате Big Data. Проанализировала данные с помощью инструментов PySpark: проверила гипотезу о равенстве баланса у мужчин и женщин; построила ml-модель, прогнозирующую отток клиентов и исследовала корреляцию и зависимость следующих признаков: кредитного рейтинга, баланса, зарплаты от возраста, пола, страны.

### Проект web-сайт для анализа данных, Февраль 2024 - Май 2024

Была Team Leadом в команде из 4 человек, занимающейся разработкой web-сайта для анализа пользовательского файла данных. На сайте представлен расчет статистик по предоставленным данным, реализовано построение графика «Ящик с усами» и графика зависимости признаков, а также построение ml-модели. Для временных рядов реализовано сглаживание 4 методами(SMA, WMA, EMA, DEMA). Используемые инструменты: Python, Pandas, NumPy, Matplotlib, Scipy, Scikit-learn, Git, SQL, PyCharm, Django.

### Проект «Определение возраста покупателей», Апрель 2023

Построила нейронную сеть с помощью модуля Tensorflow.Keras, которая позволяет определять возраст покупателей. Используемые инструменты: Python, Pandas, NumPy, Matplotlib, TensorFlow, Git.

### Проект «Восстановление золота из руды», Февраль 2023

Подготовила прототип модели машинного обучения для компании, разрабатывающей решения для эффективной работы промышленных предприятий. Обучила модель для предсказания коэффициента восстановления золота из золотосодержащей руды, что позволило оптимизировать производство, чтобы не запускать предприятие с убыточными характеристиками. Используемые инструменты: Python, Pandas, Sklearn, Matplotlib, Numpy, Joblib, Scipy, Git.

### Проект «Выбор локации для скважины», Декабрь 2022

Построила модель для добывающей компании, позволяющую определить регион, где добыча нефти принесет наибольшую прибыль. Проанализировала возможную прибыль и риски техникой Bootstrap. Используемые инструменты: Python, Pandas, Sklearn, NumPy, Scipy, Git.

## ОСНОВНОЕ ОБРАЗОВАНИЕ

**МИРЭА — Российский Технологический Университет, Институт Информационных технологий, Прикладная математика (Анализ данных) (2021 - 2025)**

## ДОПОЛНИТЕЛЬНОЕ ОБРАЗОВАНИЕ

Курс Яндекс Практикум «Специалист по Data Science». Выполнила 14 проектов: защита персональных данных клиентов, рекомендация тарифов, анализ данных интернет-магазина, прогнозирование стоимости автомобилей и др.

Курс Samsung Innovation Campus «Big Data». Выполнила лабораторные работы, используя инструменты и системы для работы с Big Data:

- работала в Hadoop, Apache Spark;
- создавала БД в Apache Cassandra;
- писала запросы к базе данных на SQL;
- работала с оболочкой Grunt и Apache Hive;
- создавала базовые функции для обработки данных и работала с RDD в PySpark;
- использовала агент Flume, чтобы собрать данные из разных источников и сохранить их в HDFS;
- использовала сервисы Kafka в командной строке для создания темы, производителей и потребителей для передачи данных через них.

## **ТЕХНИЧЕСКИЕ НАВЫКИ**

Языки программирования: Python, SQL, R, NoSQL

Инструменты разработки: Git, Jupyter Notebook, PyCharm, Google Colab

Инструменты анализа: Python - Pandas, NumPy, Scikit-learn, Seaborn, Matplotlib, Scipy, Catboost, Keras, PyTorch, TensorFlow, LightGBM, XGBoost, Statsmodels

Системы для работы с большими данными: Hadoop - HDFS, YARN, MapReduce; Apache Cassandra, Apache Pig, Apache Hive, Apache Spark, ETL: Sqoop, Flume, NiFi, Kafka

Языки: Russian (Native), English (Upper-Intermediate)