

Валерия Лепина

Data Scientist

📍 • Москва и МО

ОПЫТ РАБОТЫ

Data Engineer/DWH Analyst

Тинькофф

📅 04/2024 - Present

Стек: Jira, Greenplum, Pandas, Airflow

Обязанности:

- Проводила анализ DWH на соответствие формирования данных требованиям и форматам Центробанка
- Оптимизировала алгоритмы формирования выгрузки в БКИ и CRE и добавила новые функций, что уже позволило избежать штрафов на сумму более 30 млрд. рублей
- Реализовала систему алертинга ошибок и сбоев в отделе, что дало возможность перехватывать ошибки до их отправки в Бюро

Data Scientist

Яндекс.Маркет

📅 08/2023 - 03/2024

Участвовала в двух проектах по созданию ml моделей.

1. Разработка модели для матчинга товаров на маркетплейсе

Стек: Numpy, Pandas, Lightgbm, Catboost, Sklearn

- Реализовывала финальную часть пайплайна для бинарной классификации
- Добилась максимальной метрики $F1 = 0.92$ с помощью CatBoostClassifier

2. Реализовать векторный поиск 10 товаров с максимальным соответствием предлагаемому

Стек: Sqlalchemy, Numpy, Pandas, Sklearn, Faiss, Xgboost

- Участвовала в создании базы данных с геоименами
- Написала функции отображения схожих запросов
- На основе LaBSE обучила модель со значением accuracy = 0.86

Data Engineer

РТУ МИРЭА

📅 2020 - 2022

Стек: PostgreSQL, Tandem, Excel

Обязанности:

- Участвовала в разработке управленческой базы данных для хранения информации о 25 000 абитуриентов
- Автоматизировала несколько задач по выгрузке с помощью SQL
- Наладила систему обновления информации о данных

ОБРАЗОВАНИЕ

Прикладная математика и информатика(Математическое моделирование и вычислительная математика)

РТУ МИРЭА, Бакалавр

📅 2019 - 2023

Data Science

Яндекс Практикум, Профессиональная переподготовка

📅 2022 - 2023

КРАТКАЯ ИНФОРМАЦИЯ

Data Scientist с небольшим опытом в анализе данных, машинном обучении и визуализации.

Обладаю навыками разработки и реализации моделей предсказания, баз данных и аналитических отчетов. Успешно работала с различными инструментами для работы с данными.

Текущий опыт работы: 2 года

Предпочтения в работе: Предпочтительней удаленный или гибридный формат работы.

SKILLS

Языки

Python

C++

PostgreSQL

Инструменты

Confluence

Jira

Greenplum

Tableau

Power BI

Airflow

Docker

Kubernetes

Hadoop

Библиотеки

Pandas

NumPy

Scikit-learn

LightGBM

CatBoost

Tensorflow

КОНТАКТЫ



Почта

lutorvalerie@gmail.com



Телеграм

t.me/ObiMerlin

ПРИМЕРЫ ПРОЕКТОВ



GitHub

https://github.com/LeraLutor

LANGUAGES

Английский

B2