ALEXEY POPOV

+7(911)789-38-48 Saint-Petersburg, Russia
wibbleydock@yandex.ru Telegram Github

OBJECTIVE

Data Science / Product Analyst intern looking for internship, seeking full-time/part-time roles

ABOUT ME

Dedicated data science and statistics enthusiast who excels at performing data collection, analysis and conducting comprehensive reports. Possess great analytical skill, strong attention to details and big desire to help companies solve data-driven cases using ML/DL/Statistics.

EDUCATION

Bachelor of Computer Science, ITMO University 2022-2026
3rd year student, Software Engineering

SKILLS

Technical Skills	Python (pandas, numpy, sklearn, matplotlib, etc), A/B testing, Statistics, EDA, Machine Learning, PostgreSQL, Visualisation, Dashboards
Soft Skills	Public Speaking, Communications and Team Building
Other	Git, Linux, Docker, Bash, Shell, ELT/ETL

COURSES

Yandex SHAD Python 2024

- Numpy, Pandas, data structures, bytecode, visualisation, testing, datatime, libraries
- OOP, multiporcessing, async, design patterns, descriptors, computation acceleration, memory leaks, Tensorflow

ITMO Bonustrack Data Analysis 2023/24

- Statistics, A/B Testing, Visualisation, Classic ML

PROJECTS

Classic ML from scratch: Пишу модели классического машинного обучения "с нуля после чего проверяю корректность выполнения с помощью сравнения с библиотекой sklearn. Всё провожу на синтетических данных. На данный момент написаны kNN classification, kNN regression, linear regression, Decision Tree classification (Check here)

EDA: Analyzed and preprocessed the German credit score dataset to prepare for modeling, conducted EDA to uncover insights and patterns within the data, implemented Random Forest and Gradient Boosting algorithms to develop predictive models for credit scoring, evaluated model performance and fine-tuned parameters to achieve optimal results. (Check here)

Yandex SHAD Python tasks: Task solutions from course (Check here)

VK. Задание на стажировку: Solved a profiling assignment for VK, which involved processing time series data, performing exploratory data analysis (EDA), generating features from the dataset, hypothesis testing, and training a classification model. I chose CatBoost as the modeling approach. (Check here)