

Vagiz Daudov

Data Scientist



Email: vagiz.0282@gmail.com

About Me



Hello! My name is Vagiz. I learn and practice to be good at analyzing data, making predictions, and being a proficient coder. I like taking my time to achieve goals and show myself the results that I was not capable of sometime ago (about my cooking skills, do not mess it up:).

Favourites: **Chocolate**, **Guitar**, **Sport**, and **Data**.

Data Science projects

credits

Annotation

Analysis of credits dataset with visualizations, explanations, and machine learning estimators testing.

Credits are the main power of bank systems. Being able to separate potential overdues from disciplined returns is a crucial classification task. It is the most complicated and unusual test task which I completed as a first-step competition for a job position. Used tools: **LogisticRegression**, **RandomForestClassifier**, **CatBoostClassifier** estimators from **scikit-learn** and **catboost** Machine Learning libraries; **matplotlib** and **seaborn** libraries for data visualizations; **pandas** and **numpy** for data manipulation.

heart disease

Annotation

EDA of heart disease Kaggle dataset.

The classification task related to identification of patients with heart diseases. High level of accuracy can save people's lives. Used tools: **LogisticRegression**, **KNeighborsClassifier**, **RandomForestClassifier** from **scikit-learn**, **matplotlib**, **pandas**, **numpy**, and **seaborn**.

sampling algorithm

Project repository

Sampling dataset to fit another distribution.

The algorithm is designed specifically for certain problem - to sample dataset of issued credits such that its distribution (across some defined number of features) is approximately the same as the TTD dataset (which contains all credits, both approved (divides further into issued and not issued) and not approved). The final dataset is used to test ROC AUC score of ML models before being deployed into production pipeline. Since they are build to predict which clients will overdue credits, train and test sets must be labeled but also similar in distribution to the TTD dataset.

Python projects

students rating

Project repository

Browser automation for local purposes.

`Selenium` is the main tool used in this project. Chrome automation with `chromedriver` and html parsing with `BeautifulSoup` allow retrieving data from my University website (for privacy purposes names and links are removed from code). `matplotlib` and `seaborn` libraries visualize statistical data about student ratings.

customers cafe

Project repository

Project for managing and monitoring customers data in a cafe.

Knowledge of databases and ability to manage them are great skills in the 21st century. The project enables user to extract any relevant information about the customers. It mostly uses `sqlite3` library.

C++

console game

Game repository

C++ project for practicing with `.dll` and `boost` library.

Work Experience

Devim company

August 2022 - Current Time

DS specialist in fintech company [Devim](#). Our team solves credit risk problem and other tasks associated with service processes.

Education

HSE Saint-Petersburg

Applied Mathematics and Informatics, AADA1

2022 - 2026

For me, courses are not easy and depth of knowledge in math and programming is tremendous. Education here makes me really happy.

Courses

There are certificates of courses which I have completed to diverse my knowledge:

Probability Theory

Basic C++

Other proficiencies

Alongside with Univesity studies and self-development I have the following skills and certificates:

- IELTS 7.5 - C1 level of English.
- French B1.
- Waterpolo Candidate in Master of sports.

Vagiz Daudov - vagiz.0282@gmail.com - References on request