(+31) 687114011
Amsterdam, Netherlands
margo.razumeyeva@gmail.com

# Margo Razumeyeva

Data Scientist

github.com/razmarrus
linkedin.com/in/razmarrus

As a dedicated Data Scientist with a strong focus on Machine Learning, I specialize in models development and extraction of insights for customer and financial analytics. Particularly in the areas of Sales prediction and Optimization of Marketing Strategies for Big Pharma. My experience in full-stack development allows me to take a holistic approach to problem-solving.
I am authorized to work in the Netherlands and am currently based in Amsterdam.

## Skills

| | |
|---|---|
| **ML Approaches** | Gradient Boosting, Linear Models, Clustering, CNN, RNN, Reinforcement learning. |
| **Advanced** | Python, GIT, Pandas, SKLearn, Catboost, LightGBM. SciPy, NumPy, MatPlotLib. |
| **Intermediate** | Keras, Tensorflow, SQL, AWS, Google Cloud, Airflow, CMD, Bash. |
| **Basic** | lua (LOVE), Kotlin, C#, ASP .NET, Xamarine, Android, Maple. |

## Experience

**SBDA Group - IT consulting** **Nov 2021 — Present**
*Data Scientist* *Remote*

End-to-end model development, ensuring seamless execution of the entire process. From data collection and preparation to model creation, deployment, and updated automation. I collaborate with clients and colleagues at every step of the process, delivering regular presentations.

**Clustering of customers or geo units. Obtained from 3 to 8 business-interpretable clusters.**
Development and integration into GBQ from scratch for several clients. Based on the data, I chose the best algorithm (K-means, DBScan, Birch) and combine its output with additional analysis (ABC, RFM).

**Data quality check and statistics package.**
Tool for data quality checks and core statistic calculation: cleaning, typing, filtering, outliers, correlation, statistic tests, Levenshtein distance, etc. I was focused on efficiency improvements with paralleling, refactoring, and class encapsulation.

**Recommendation system for omnichannel communication. Open-rate increment by 8%.**
Development and update automation of a set of Gradient Boosing models. Each model was trained and tuned separately on a dataset tailored to each customer communication channel, including emails, call-center calls, and offline events.

**Sales prediction. The interpretation of model weights.** $WAPE = 0.06, R^2 = 0.98$
Various linear models were trained until giving a solid sales prediction. Different data preparation techniques and feature complications were researched. Best model (Elastic net) was giving prediction for next 4 month.

**Evaluation of additional profit made with promotion channel.**
Creation of a model from scratch and deployment into the Airflow.
The boosting model was trained to predict sales based on customers' history of interaction with content and historical sales data. Then, through analysis of changes in model prediction, each promotion channel was valued in terms of additional profit.

## Education

**Bachelor of Computer Science**, *Belarusian State University of Informatics and Radioelectronics* Sep 2017-Jun 2021
*Majors: Computer science (70 credits), Mathematical engineering (85 credits), Full-stack software development (60 credits).*
*Minors: Social Sciences (45 credits), Engineering (25 credits)*

## Pet Projects

### Forecasting the Exchange Rate — April 2021

Several supervised learning models that solve the problem of predicting financial time series like Gradient Boosting, Random Forest, SVM, Linear regression. Project made from scratch with pandas and scikit-learn.

### Reinforcement learning in Dino Run — May 2019

Reinforcement learning (Q-Learning) with the usage of CNN for the Dino Run game. The RL-based agent takes data from CNN that processing game screenshots. Two N N made with Keras, NumPy, and a raw python interlayer to connect them.

### Convex Optimization with NumPy — May 2020

Pack of tasks on Convex Optimization focused on simplex methods, quadratic programming, and graph theory. Solved with NumPy.