

KOREPANOV ALEXEY

Data Scientist

@ grazder@yandex.ru
github.com/grazder

+7 (919) 915-43-30
kaggle.com/grazder

Saint Petersburg, Russia
@grazder



PROJECTS

OK Caption System

Odnoklassniki Study Project

Spring 2021

- We developed an application for voice assistant (Marusia) which reads news feed from Odnoklassniki and generates image captioning.
- I tried different LSTM and Transformer Image Captions SOTA models based on ResNet and Faster R-CNN features during the project.

Image Captioning Computer Vision NLP PyTorch API

COMPETITIONS

IEEE-CIS Fraud Detection

Kaggle

Fall 2019

- The aim was to detect fraud from customer transactions. Binary classification problem with a large number of anonymized features.
- During the competition I tried and applied a lot of different ML techniques like Up/Down Sampling, Adversarial Validation, Blending, Stacking. I also learned different Feature Selection, Validation, and Parameter Tuning approaches. I experimented a lot with PCA, XGBoost, LightGBM, and Catboost. All computations were made in the Google Cloud.
- As a result, my best single CatBoost model achieved a Top-50 score on the final leaderboard. This model was the best model in our team's final stacking. My team took 23rd place on the final leaderboard with that stacking. You can find a code in my [GitHub repository](#).

Feature Engineering LightGBM CatBoost PCA Google Cloud

Yandex Cup 2019

Yandex

Fall 2019

- In the qualification round I solved many problems on different topics such as Algorithms, Machine Learning, Probability Theory.
- In the final round, the task was to classify texts by topic. A dataset with text was given, and the challenge was to understand what topics it relates to.
- I solved this task using CountVectorizer, TfidfVectorizer, and Logistic Regression. I selected topics by the threshold.
- I tuned this model a lot and achieved 6th place on the private dataset.

Machine Learning NLP CatBoost Algorithms and Data Structures

Profi.ru Hackathon

Skoltech & Profi.ru

Feb 2021

- The aim was to identify users by their typing characteristics to prevent fraud.
- We made a LightGBM model with various features. Also applied Adversarial Validation while collecting new data from raw users dataset.

Machine Learning LightGBM Feature Engineering

SKILLS

Python Machine Learning
Deep Learning PyTorch
XGBoost LightGBM CatBoost
Sklearn Git Linux Bash
Docker OpenCV C# Django
Algorithms and Data Structures

EDUCATION

Master's program Machine Learning and Data Analysis

NRU "Higher School of Economics"

Saint-Petersburg Sept 2020 – Present

Open Machine Learning Course mlcourse.ai

ods.ai Rank: 6 / 1600

Spring 2019

Bachelor's program Software Engineering and Information Systems Management

The Glazov Korolenko State Pedagogical Institute

Glazov Feb 2018 – Jun 2020

Bachelor's program Software Engineering

NRU "Higher School of Economics", Moscow

Moscow Sep 2016 – Dec 2017

COURSEWORK

Machine Learning Deep Learning
Reinforcement Learning
Machine Learning on Big Data
Optimization Methods
Numerical Methods in Linear Algebra
Computer Vision Probability Theory
Mathematical Statistics