ARTEM RYBLOV

+7(952) 44-98-154
ryblovartem@gmail.com ◇ LinkedIn ◇ GitHub

ABOUT

I'm a skilled Data Scientist with more than 4 years of experience in data science. I've helped companies optimize their current AI solutions and implement new ones.

I get excited about new opportunities where I can apply my skills to create solutions that help businesses achieve their goals and improve the lives of end users.

SKILLS

Programming	Python, SQL
DS Frameworks	Pandas, PySpark, NumPy/SciPy, Matplotlib/Seaborn, BeautifulSoup, Loguru
ML/DL/NLP Frameworks	Scikit-learn, Catboost/XGBoost/LightGBM, Optuna, PyTorch, Hugging Face, Sentence Transformers, NLTK, Gensim, Spacy
Tools/Platforms	Git, Docker, Jira

EXPERIENCE

Data Scientist Jul 2022 - Present
OneFactor *Remote*

- Using telecom data did:
  - 3x Credit Scoring models
  - 1x Anti-fraud model
  - 3x Lead Generation models
  - 5x AutoML models
  - 6x Geoanalytics reports
  - 8x Ad-hoc tasks
- Using the principles of clean code:
  - Refactored >15 jupyter notebooks
  - Transformed set of raw notebooks into Geoanalytics framework
- Implemented AB testing for model comparison using McNemar test

NLP Engineer May 2019 - Feb 2022
HARMAN *Nizhny Novgorod, Russia*

- Developed an end-to-end Offline NLP framework for training and testing Intent/Token Classification (NER) models for mobile devices (tf-lite), which was successfully applied for two customers in 2021.
- Implemented semantic search approach (pretrained embeddings + approximate nearest neighbours) for intent classification that significantly improved the quality of predictions, training & inference time, and was accepted in production as the main intent classifier for voice assistant.
- Led the NLP team and took part in road-map planning, defining the main directions of development and detailing them.
- Took part in the development of personal cruise assistant MSC Zoe (NLP):
  - Developed Noise Sentences Classifier to filter out truncated sentences which improved UX;

  - Enhanced Entity Linking with Similarity Algorithm based on the Levenshtein distance which led to better recognition of named entities.

PROJECTS

Shows Analysis
I've scraped information about top-1000 movies and top-1000 series, and started investigation.

- Parsing web-pages
- Data Analytics:
  - Research on the quality of localization of movie titles
- Data Science
  - Sentiment Analysis of Show Reviews

Small Projects

- Churn Prediction | Binary Classification, EDA, Catboost + Optuna
- Salary prediction | Regression, DL, NLP, EDA, PyTorch
- Simpsons classification | Multiclass classification, DL, CV, Transfer Learning, PyTorch

EDUCATION

Master of Economics, Higher School of Economics 2016 - 2018
Cum. GPA: 9.0 / 10.0

Bachelor of Computer Science, State University of Nizhny Novgorod 2012 - 2016
Cum. GPA: 4.5 / 5.0

PUBLICATIONS

- Comparison of Machine Learning Methods for Analysis of Ulcerative Colitis Proteomic Data
- Parenclitic Network Analysis of Methylation Data for Cancer Identification