

# ARTEM RYBLOV

[ryblovartem@gmail.com](mailto:ryblovartem@gmail.com)

[linkedin.com/artemryblov](https://linkedin.com/artemryblov) ♦ [github.com/Extremesarova](https://github.com/Extremesarova)

## EXPERIENCE

---

### Senior Data Scientist

Jul 2022 - Present

OneFactor

*Remote*

- Built anti-fraud model, 3x lead generation models, 7x credit scoring models, 8x geoanalytics reports on telecom data and 11x AutoML models with estimated yearly impact on revenue at around €1.3m
- Improved the process of model selection for main scoring pipeline by implementing A/B Test (DeLong test)
- Transformed set of raw notebooks into geoanalytics framework and refactored > 15 Jupyter Notebooks using the principles of clean code

### Data Scientist

May 2019 - Feb 2022

HARMAN

*Remote*

- Developed an end-to-end Offline NLP framework for training and testing Intent/Token Classification models for mobile devices (tf-lite), which was successfully applied for two customers in 2021 and generated more than €0.2 m in revenue.
- Implemented semantic search approach (pretrained embeddings + approximate nearest neighbours) for intent classification that significantly improved the quality of predictions (10%), training (250x) & inference (5x) time, and was accepted in production as the main intent classifier for voice assistant.
- Led the NLP team and took, defined and detailed the main directions of development.
- Took part in the development of personal cruise assistant MSC Zoe:
  - Developed Noise Sentences Classifier to filter out truncated sentences which improved intent classification by 15%;
  - Enhanced Entity Linking with Similarity Algorithm based on the Levenshtein distance, which led to 40% better recognition of named entities.

## SKILLS

---

### Programming

Python, SQL

### General Frameworks

Pandas, PySpark, FastAPI, Matplotlib/Seaborn, BeautifulSoup, Loguru

### ML Frameworks

Scikit-learn, Catboost/XGBoost/LightGBM, Optuna, PyTorch,

Hugging Face, Sentence Transformers, NLTK, Gensim, Spacy

### Tools

Git, Docker, Jira

## PROJECTS

---

### Shows Analysis

- [Parsed 206737 reviews and information about 1962 shows](#) | BeautifulSoup
- [Built baseline model for sentiment classification \(0.8 F1 score\)](#) | Binary Classification, Logistic Regression, TF-IDF
- [Deployed Sentiment Classifier](#) | Python, Docker, FastAPI

### 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

---

|  |             |
|--|-------------|
| <b>Bachelor of Computer Science</b> , State University of Nizhny Novgorod<br>Cum. GPA: 4.5 / 5.0 | 2012 - 2016 |
| <b>Master of Economics</b> , Higher School of Economics<br>Cum. GPA: 9.0 / 10.0                  | 2016 - 2018 |

## PUBLICATIONS

---

- [Comparison of Machine Learning Methods for Analysis of Ulcerative Colitis Proteomic Data](#)
- [Parenclitic Network Analysis of Methylation Data for Cancer Identification](#)

## PERSONAL DEVELOPMENT

---

### [The Pillars of Data Science](#)

I've created a site on Notion where I have been developing two differently styled roadmaps for learning Data Science based on the links I share on this [telegram channel](#).

Both guides contain the same information but are formatted differently for your convenience.

The first roadmap is called **Topic Guides**. These guides are focused on topics like Machine Learning and then split into knowledge levels and resource types. Thus, you can use them if you want to focus on a specific topic and deepen your knowledge.

The second roadmap is called **Content Type Guides**. These guides are targeted by resource type, such as Courses, and then divided into topics and knowledge levels. So you can use them if you prefer a certain type of resource and want to expand your knowledge.

### **Blog articles:**

- [Research on the quality of localization of movie titles](#)
- [\(Not IMDb\) Movie Reviews Dataset EDA](#)
- [Text Classification: Baseline with TF-IDF and Logistic Regression](#)