

# Roman Kutubidze

## Data Scientist

Analytically minded self-starter with a 4 plus years of experience collaborating with cross-functional teams and ensuring the accuracy and integrity around data and actionable insights

r.kutubidze.ai@gmail.com

[LinkedIn](#)

[github](#)

[Personal Website](#)

## Work Experience

### WTX – Machine Learning Engineer

September 2022 - current

Lisbon, Portugal

- Developed "on-shelf" model to classify all products in stock based on internal and external data
- Trained and deployed model in production using Azure machine learning studio. I used github actions for CI/CD, Terraform for the infrastructure and Mlflow as an orchestrator
- Architected and designed efficient and scalable data pipelines that can extract, transform, and load data from several sources into the data warehouse or data lake. For example ETL than extracted data from Hubspot using Hubspot api, transform and loaded into Azure data lake
- Data modeling and schema design: designed and implemented database schemas, defined tables, fields, relationships, and constraints to support efficient data storage and retrieval

### EPAM – Data Scientist

January 2022 - October 2022

Tbilisi, Georgia

- Named entity recognition model for information retrieval project
- Fine-tuned BERT language model using Python and huggingface and implemented a two-stage transfer learning pipeline
- Performed code reviews, optimized for data processing efficiency and integration
- Wrote and developed production-level code/pipelines
- Collaborated with DevOps engineers and Data engineers to bring models to production

### Sweeft Digital – Software Engineer, Part-time

May 2021 - December 2021

Tbilisi, Georgia

- Worked on web apis using Django and Django rest

### TBC – Data Scientist

September 2019 - December 2022

Tbilisi, Georgia

- initiated and developed income estimation for non-payroll clients based on transactions and demographic data using Python, Tensorflow, SQL, Pyspark, Mlflow
- Default credit score model for PAL loans. predicts if customer goes default or not in some time window. Communicated a lot with risks department about lots of hard rules and different business limitations specifically for this project
- Developed customer churn prediction model using scikit-learn
- Sentiment analysis - This project was about to create sentiment analysis model for clients's feedback using BERT, Mlflow, Pyspark and SQL

## Education

### University of L'Aquila – Master's, Condensed Matter Physics

September 2018 - June 2019

L'Aquila, Italy

## Skills

- Python: Tensorflow/Pytorch, NLP : transformers/BERT/gpt 2; Web: Django/Flesk; ML/Data Engineering: AWS/Azure, Mlflow, Airflow, github actions