Roman Kutubidze

Data Scientist

r.kutubidze.ai@gmail.com
[LinkedIn]
[github]
[Personal Website]

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

Work Experience

Thaleria – Machine Learning Engineer

July 2023 - current   Lisbon, Portugal

- Designed and developed MLOps solution for the fastapi application with Elasticsearch on Azure
- Developed CI with Azure Devops yaml pipelines
- Designed and implemented Azure ML pipelines to train and predict the model
- Collaborated with the AI team from the client's side on the project's Architecture and implementation details
- Designed and implemented unit tests

WTX – Machine Learning Engineer

September 2022 - June 2023   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

June 2022 - October 2022   Tbilisi, Georgia

- Named entity recognition model for information retrieval project
- Performed code reviews, optimized for data processing efficiency

Sweeft Digital – Software Engineer, Part-time

Aug 2021 - December 2021   Tbilisi, Georgia

- Worked on web apis using Django and Django rest

TBC – Data Scientist

September 2019 - May 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

Education

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

Skills

- Python, Pandas, Numpy, Tensorflow, Pytorch, NLP , Transformers, Huggingface, LangChain, Pyspark, Django, Flesk, Docker, Bash, SQL, AWS, Azure, Mlflow, Airflow, github actions, Terraform, Algorithms and data structures, CI/CD, ETL