TOLGA ŞAKAR
Data Scientist

Contact Information

Email: tolgasa2@gmail.com Github: github.com/dfavenfre

Phone: +90 539 772 13 42 Kaggle: kaggle.com/dfavenfre/code

Medium: medium.com/@bauglir LinkedIn: linkedin.com/in/tolga-şakar-575b86136

Programming Languages & Skills

- Python (Pandas, Numpy, Statsmodel, Scikit-learn, Tensorflow, Keras, PyTorch, LangChain, LangGraph), Julia (Flux, Zygote),
- SQL
- Development Frameworks (Docker, FastAPI, Streamlit, HTML, CSS)
- LangChain (Conditional Tool Calling, Customized Tool Architecture, Few-shot Prompt Engineering, Chatbot Security)
- LangGraph (Customized Multi-Tool Agent Architecture, Flow Engineering with multi-tools, Chatbots with Human-Feedback)
- Transfer Learning (TensorflowHub, HuggingFace)
- Predictive Modelling (Time-Series Forecasting, CARTs)
- Optimization (HyOPT, Keras Pruner, Optuna, Randomized Search, Grid Search Bayesian)
- ML models, LLMs and Chatbot Development and Monitoring (Weights & Biases, LangSmith, Predibase, OpenAI Fine-tuning)
- Web Scraping (Selenium, PlayWright)

Projects

- **TalkYou**

  TalkYou is an innovative open-source project designed to enable users to have a chat with any YouTube video. It brings user a customized chatbot experience, not only with the ability to chat but also with an amazing feature of image retrieval based on user queries. Both LangChain and LangGraph frameworks were utilized as the back-bone of this project to achieve conditional tool calling capabilities.

- **LLMRoboFund**

  LLMRoboFund is a powerful financial assistant chatbot trained with TEFAS documents and database. The chatbot is equipped with a customized multi-tool agent (Text-to-SQL tool and RAG tool) , aiming to ease the investment research process. Chatbot development had undergone various optimization steps, such as RAG optimization with various vectorstores, RAG methods, embedding models (multilingual) to reduce token usage, speed up response time and increase response accuracy.

- **Econ-Dashboard (Multi-Model Streamlit Application)**

  Created a centralized financial dashboard on Streamlit for screening financial/economic data, sentiment classification, and time-series forecasting. This all-in-one dashboard is equipped with hand-crafted neural network models, including a sentiment classifier with pre-trained sentence

transformer embedding, and LSTM time-series forecasting models, from scratch, distinguished for different market capitalizations.

- **Bank Customer Deposit Classifier (Churn Prediction)**

A Streamlit app is developed upon a classification model that is trained on a Portuguese banking institution data to make predictions whether a bank customer would subscribe to a term deposit. Various machine learning techniques were used throughout model training, such as SMOTE to solve imbalanced-label issue, RFECV to solve dimensionality issue by eliminating features with less contribution.

- **Electricity Price Forecasting Modelling**

Developed various predictive models, including a hand-crafted neural network (LSTM) time-series model and an XGB regression model to predict t+1 hourly electricity prices of the EXIST Market Transparency Platform.

Work Experience

Data Scientist

Naviga AI (Ankara,TR)

Dates: 01/2024 – Present

I am responsible for developing chatbots utilizing the LangChain and LangGraph frameworks to create customized and multi-tasking tools and agents. I take direct responsibility for chatbot development processes, such as RAG optimization, developing customized tools and agents, and prompt engineering to both instruct large language models and to prevent chatbot injection attacks.

I have hands-on experience working with LLMs, Speech-to-Text, Text-to-Speech models, Multi-Modal RAG, and fine-tuning LLMs with few-shot examples. Additionally, I engage in back-end development utilizing the FastAPI framework to provide endpoints for chatbots.

Data Scientist

Bosphorus AI (Ankara,TR)

Dates: 10/2023 – 01/2024

Achievements/Tasks

I was involved in a variety of projects, including the development of a fraud detection model for the EXIT Platform. This model aims to identify market participants engaged in price manipulation within order books. In addition, I was tasked with creating chatbots that guide customers toward recommended insurance products for an insurance company. Furthermore, I was also responsible for developing recommendation models based on customer segmentation tasks.

Equity Research Analyst

Valens Research (Istanbul, TR)

Dates: 01/2022 – 03/2023

Achievements/Tasks

My responsibilities encompassed the preparing of investment recommendation reports achieved through a sequence of rigorous data-driven analyses. This included intricate financial and economic data modeling, as well as thorough peer comparisons. I frequently used SPSS IBM for statistical inference and Python when conducting time-series analysis.

Quantitative Research Intern

E2T (Remote, UK)

Dates: 07/2021-10/2021

Achievements/Tasks

I worked in quantitative research projects, acquiring methodologies to derive insights using statistical inference techniques, revealing trends in finance and economics.

Education

TED University

Master of Science in Applied Data Science

GPA: 3.62 / 4

Key Courses

- Machine Learning
- Deep Learning for Data Science
- Information Retrieval
- Modelling and Analysis of Uncertainty
- Time-Series Forecasting
- Numerical Methods
- Prompt Engineering

TED University

Bachelors of Science in Economics & Business Administration

GPA : 3.3 / 4

Certificates / Udemy Courses

- TensorFlow Developer Certificate - Udemy
- PyTorch Developer Certificate - Udemy
- Feature Selection for Machine Learning - Udemy
- Hyperparameter Optimization for Machine Learning - Udemy
- LangChain with Python Bootcamp - Udemy

- Docker Masterclass for Machine Learning and Data Science - Udemy
- Python Object Oriented Programming (OOP) - Udemy
- Unsupervised Learning (K-means, KNN, PCA) - DataCamp
- Supervised Learning (CARTs) – DataCamp