Soroosh Sorkhani

soroosh.sorkhani@gmail.com

sorooshsorkhani.github.io

sorooshsorkhani Google Scholar soroosh-sorkhani Toronto

Summary

- Data Scientist with 6+ years of experience in machine learning, LLMs, and computer vision. Skilled in developing and aligning ML solutions to business needs. Proven track record of building and presenting high-impact models for diverse industries.

Skills

- **Programming**: Python, R, SQL, NoSQL
- **Libraries & Frameworks**: LangChain, HuggingFace, TensorFlow, Pytorch, scikit-learn, L2R, Pandas, NumPy
- **ML & AI**: LLMs, Computer Vision, NLP, Deep Learning, Time Series, Graph Analysis
- **Systems & tools**: Unix/Linux, NVIDIA Jetson Nano, AWS

Experience

- **GEO BON - McGill University** — Montreal, Canada
  Data Scientist, AI Engineer — Sep 2024 - Present
  Developed an AI assistant to assist 3,500+ members of GEO BON, a global network of biodiversity researchers spanning 152 countries and 2,000 organizations. The **RAG-powered agent** generates responses by leveraging open-source LLMs such as **Llama** and **DeepSeek**, offering capabilities ranging from coding and statistical data analysis to delivering reference-backed, reliable, and actionable solutions for real-world applications.

- **York University** — Toronto, Canada
  Data Scientist, Computer Vision Engineer — Jun 2023 - Sep 2024
  Developed computer vision systems to address urban mobility challenges. Built a **Smart City Signals** system using **YOLOv8** for real-time traffic analysis at intersections, achieving **0 error** in lane-wise vehicle and pedestrian detection while estimating speed and waiting times for deeper insights. Developed a parking lot occupancy detection system, utilizing YOLOv8 for vehicle detection. Led the design and deployment of an on-street parking detection system, utilizing GPS data and dashboard camera feeds to identify parked vehicles, enhancing parking management and enforcement strategies. Additionally, implemented machine learning models, including XGBoost, to predict parking violations across a network of 940 parking locations in Downtown Toronto. The model effectively predicted the number of violations for given location-time instances, with a **Mean Absolute Error (MAE) of 1.3**. *(See a live demo on my website)*

- **Canadian Tire Corporation** — Toronto, Canada
  Data Analyst — Sep 2021 - Sep 2022
  Managed and optimized databases using SQL, creating reports and dashboards with actionable insights for stakeholders. Performed statistical analyses, including **A/B testing**, to compare two different transportation methods. Led the enhancement of a data-driven report encompassing over 5000 daily records on transportation equipment repairs. Achieved operational efficiency by integrating data from multiple databases and automating processes using Python, reducing daily processing time by 50%.

- **Toronto Transit Commission (TTC)** — Toronto, Canada
  Machine Learning Consultant — Jan 2021 - Jun 2021
  Assessed the end-to-end process of a machine learning product for TTC to improve real-time passenger count predictions at subway stations. Conducted a thorough review of the product pipeline, identifying issues in data collection, validation, and evaluation. Provided actionable recommendations to optimize data sourcing, model development, and evaluation processes. Proposed four potential solutions, including feature engineering and advanced time series models such as RNN and LSTM, to improve the overall performance and reliability of the model.

- **Toronto Metropolitan University** — Toronto, Canada
  Data Scientist — Sep 2019 - Jun 2021
  Built an expert **recommendation system** for Q&A platforms, including Stack Overflow. Engineered 74 features using methods like **LDA topic modeling**, **text similarity**, and **graph embedding**. Proposed a **random forest learning to rank** approach that achieved a 16.41% improvement in NDCG@10 over the state-of-the-art model across datasets from five websites, with over 15,000 users and 21,000 questions. Analyzed key features to understand their impact on model performance.

Education

- **DeepLearning.AI** — Online
  Deep Learning Specialization on Coursera — Jun 2022 - May 2023
  Courses: Neural Networks and Deep Learning, Improving Deep Neural Networks: Hyperparameter Tuning, Regularization and Optimization, Structuring Machine Learning Projects, Convolutional Neural Networks, Sequence Models
- **Toronto Metropolitan University** — Toronto, Canada
  Master of Science in Electrical and Computer Engineering — Sep 2019 - Jun 2021
  Thesis: Feature-based Question Routing in Community Question Answering Platforms

Publications

- Sorkhani, S., Etemadi, R., Bigdeli, A., Zihayat, M., & Bagheri, E. (2022). Feature-based Question Routing in Community Question Answering Platforms. *Information Sciences*