Dhruthick Gowda Mohan

San Diego, California, 92092

858-323-7769 dgmohan@ucsd.edu linkedin.com/in/dhruthick github.com/dhruthick dhruthick.github.io

Summary

UC San Diego Graduate with two years of professional experience in Data Science and Machine Learning a Master's Degree in Computer Science. I am actively seeking full time opportunities starting in July 2024. My interests lie in Machine Learning, MLOps, Data Science, and Data Analytics.

Education

University of California, San Diego **Sep. 2022 – Jun. 2024**

Master of Science in Computer Science, GPA: 3.97/4 *La Jolla, California*

- *Coursework*: Data Systems for ML, Big Data Analytics, Search & Optimization, Probabilistic Learning, Trustworthy ML, Database Systems, Numerical Linear Algebra, Computer Architecture
- *Teaching Assistant*: Statistical NLP, Recommender Systems & Web Mining, Convex Optimization Algorithms
- Helped organize Carpentries Workshops for Python and Data Management at the School of Global Policy & Strategy.

Ramaiah Institute of Technology **Aug. 2017 – Jul. 2021**

Bachelor of Engineering in Information Science (Honors), GPA: 9.61/10 *Bengaluru, India*

- *Coursework*: Distributed Computing, Operating Systems, Data Structures, Algorithms, Distributed Computing

Technical Skills

Languages: Python, SQL, Java, Go, R, C, C++, Javascript, HTML
Development Tools: Spark, AWS (EC2, ECS, S3, Lambda, SageMaker), GCP BigQuery, Docker, Kubernetes, MySQL, MongoDB, Jupyter Notebook, Bash, MinIO, AirFlow, Flink, Superset, Git, GitHub
Frameworks: PySpark, TensorFlow, PyTorch, SciKit-Learn, Pandas, SciPy, Numpy, Flask, Node.js, HuggingFace

Industry Experience

Data Scientist **Aug. 2021 – Aug. 2022**

India Urban Data Exchange (IUDX) *Bengaluru, India*

- Led the development of a road network construction algorithm, resulting in a 15% increase in traffic modeling accuracy and a 32% reduction in inference time, leading to more efficient visualization of the transit management data. This innovation is documented in a preprint (link)
- Analyzed petabytes of geospatial smart city data from diverse domains including transit management, air quality, and emergency services, leveraging efficient visualization techniques with exploratory data analysis and ETL operations. This analysis led to the identification of critical insights that were instrumental in enhancing urban infrastructure.
- Applied statistical modeling, advanced machine learning (supervised and unsupervised) and deep learning algorithms, including regression, clustering, CNNs, and LSTMs, to solve complex problems such as correlating vehicle traffic with air quality and developing predictive traffic models. These efforts were found to be promising in improving urban planning and environmental sustainability.
- Developed and deployed a comprehensive dashboard showcasing smart city data insights to key stakeholders, including the Ministry of Housing and Urban Affairs. This dashboard highlighted the potential of leveraging data across city departments for enhanced collaboration and efficiency gains with data-driven decision-making.

Data Science Research Intern **Jan. 2021 – Jul. 2021**

India Urban Data Exchange (IUDX) *Bengaluru, India*

- Enhanced a custom Python SDK to retrieve both historical and real-time smart city data, resulting in a 60% reduction in data retrieval time. Facilitated company-wide usage with proper authorization protocols.
- Conducted an in-depth analysis of air quality sensor data across the city, utilizing geospatial interpolation techniques for effective visualization. Identified key areas that can enhance air quality monitoring coverage over the city.
- Developed a multi-class classification model using recurrent neural networks (RNNs) to categorize grievances submitted by citizens in Gujarati, translated to English. The method reduced processing time for citizen grievances, potentially enabling efficient response.

Research Experience

Graduate Student Researcher **Apr. 2024 – present**

Center for Applied Internet Data Analysis (CAIDA) *San Diego, California*

- Refined a distributed system using AWS Lambda for querying Ookla speed test servers, optimizing request handling to mitigate IP blocking and ensure reliable data collection with dynamic querying and load distribution strategies, enhancing efficiency and performance.
- Analyzed and visualized Ookla Open Data, examining global internet performance metrics and server data trends.
- Developed a Go script utilizing Go threads to efficiently query network measurement data through GCP BigQuery, collecting millions of traceroute records from various ISPs for research analysis.

Projects

Lyric-Based Playlist Continuation with Language Models | *Python, PyTorch, LightFM, XGBoost* **Jun. 2023**

- Studied the effectiveness of lyric-based features in music recommendation with mood classification of song lyrics using BERT, and recommendation of tracks using a two-stage approach with matrix factorization and XGBoost.
- The BERT model achieved an accuracy of 94% and the recommendation model achieved an R-Precision@500 of 0.5667.

Neural Collaborative Filtering for Automatic Playlist Continuation | *Python, TensorFlow* **Dec. 2022**

- Developed a recommender system with neural collaborative filtering that leverages the Spotify Million Playlist Dataset.
- Achieved a notable NDCG@10 score of 0.705, surpassing the performance of Singular Value Decomposition.
- Incorporated playlist-track interactions and track meta data to deliver ranked recommendations.

Instruction Branch Prediction with Perceptron | C **Mar. 2024**

- Utilized machine learning techniques to improve instruction branch prediction accuracy against existing approaches - GShare and Tournament.
- Achieved a 25% gain in prediction accuracy, leading to enhanced performance.

Trend Detection in Stock Prices | *Python, Flask, TensorFlow, SciKit* **Jun. 2021**

- A web application that scrapes real time stock market data, detects the current trend (Uptrend/Downtrend) of a given stock and attempts to predict the next day's closing price. Random Forest and an LSTM, with data from India's NIFTY 50 market, were used to develop the detection and prediction model.

Brain Tumor Detection and Segmentation in MRI with Object Detection | *Python, TensorFlow* **Jan. 2021**

- Constructed a deep learning model that uses the VGG16 and Mask R-CNN architectures with transfer learning to detect the presence of tumors in MRI images and locates them through pixel-based segmentation.
- Classification accuracy of 90% with an IoU score of 82% was achieved. (paper link) (pdf)

Network Intrusion Detection | *Python, SciPy, SciKit-Learn* **Jul. 2020**

- Built a machine learning model that detects malicious attacks in network traffic of a system and further classifies the type of attack. Various machine learning algorithms were tried before arriving at a Random Forest-based model that gave an accuracy of 98%.

Publications

- Shivika Sharma, Nandini Mawane, Dhruthick Gowda M, Mayur Taware, Chetan Kumar, Yash Chandrashekhar Dixit, and Rakshit Ramesh (2023). Estimating time of arrival of vehicle fleets with GCN based traffic prediction. (preprint)
- Dhruthick Gowda M, Neralakatte Prajwal Pai, Anirudh R, Shruthi G, and Dr. Krishna Raj P M (2020). Brain Tumor Detection and Segmentation Using VGG16 and Mask R-CNN with Transfer Learning. Solid State Technology, Volume 63, Issue 5. (link) (pdf)