Harshita Vemula

harshitavrao@gmail.com | Website | 940-205-6305

WORK EXPERIENCE

Applied Scientist, Flyrlabs Jan 2023 − July 2024

- Collected and analyzed cargo data, providing actionable insights through statistical analysis. Developed Looker dashboards and conducted exploratory data analysis to identify patterns and trends. Designed schemas and wrote queries to generate production-level data tables. Built data processing pipelines using Dataflow and Beam in Vertex AI. Collaborated directly with clients and cross-functional teams to understand requirements, present results, and iterate based on client feedback.
- **Forecasting**: Developed and benchmarked traditional time series models (ARIMA, MA etc.) alongside advanced multi-headed deep learning models (LSTM, RNN, transformers, NHITS, NBEATS) to accurately forecast revenue and bookings. Conducted extensive experimentation with various model layers and architectures to optimize performance, incorporating both static and dynamic data inputs for improved accuracy and robustness.
- **Efficiency Improvement and Pricing Model Development**: Successfully reduced model training time by 40% through optimizing feature engineering in the codebase. Currently developing a sophisticated pricing model to predict customers' willingness to pay for flight tickets, leveraging RL techniques.

Data Scientist, Revenue Science, Twitter July 2021 − Oct 2022

- **Brand awareness team** - Enhanced the advertiser experience on Twitter by building and pricing products effectively.
  -Campaign planner: Developed ML models to forecast reach, frequency, and CPM for various ad formats, aiding in campaign planning. This tool was adopted by 60% of advertisers. Defined KPIs and created dashboards to monitor the planner's performance. Additionally, built tools to predict advertiser underspending patterns and recommended counter-strategies.
  -Ad product pricing: Collaborated with sales and ad auctioning teams to conduct data-driven pricing analyses. Formulated pricing strategies for the Takeover ad product, which generates over $300M annually.
- **Ads privacy team** - Strategized serving win-back prompts to users for in-app privacy settings by performing observational studies; Conducted revenue and signal impact analysis of new privacy policies to help drive action strategies and resource allocations.

Data Science Intern, Dell May 2019 − Oct 2019

- Built clustering based ML model for anomaly detection to identify different operating states of a server from the telemetry data obtained; Feature, expected to have a revenue potential of $200M per year, is to be provided as a service to be used for intrusion detection, optimal scheduling of jobs etc.
- Implemented Mean Shift Clustering algorithm to define a baseline; augmented data using SMOTE and trained an Autoencoder to obtain reconstruction error for novelty detection.

Student Consultant, Affinity Answers Jan 2019 − May 2019

- Analyzed purchasing patterns of the client and built a collaborative filtering based recommender system to suggest new audience segments.
- Heuristically identified high-performing audience segments and clusterized matrix factors to extract similar segments; calculated trends, seasonality of the new segments suggested to be added to the shelf using google trends data. Boosted the sell through rate of the segments on shelf by 1.5 times.

EDUCATION

The University of Texas at Austin May 2021
M.S. in Operations Research and Industrial Engineering.

ACADEMIC PROJECTS

Reinforcement Learning for efficient energy management. Oct 2019 − Dec 2019

- Implemented reinforcement learning (RL) algorithms to develop an agent for demand response management aimed at minimizing the overall energy consumption cost from the grid. Defined the state space, action space, and reward functions for the system. Explored temporal difference learning methods like Q-learning and SARSA.

ACHIEVEMENTS

Microsoft Azure Hackathon - Winner Apr 2019

- Won the 2019 Data Hack Challenge organized by UT MLDS group and Microsoft azure and Oracle cloud. The competition saw partition from over 100 teams from UT Austin.
- Forecasted rental demand at kiosks of a bicycle sharing company using Random Forests; predicted hourly position of 5000 bicycles using stochastic modelling. Suggested a dynamic pricing framework for the network and proposed a relocation strategy.

TECHNICAL SKILLS

Software tools: GCP, Python, SQL, Tensorflow, Numpy, Pandas, Scikit-learn, Docker, Git, Vertex AI, Airflow, Spark, Beam, Looker, Wandb

Machine Learning: Classification, Regression, Clustering, Anomaly detection, Dimensionality reduction, Deep Learning, Reinforcement learning, Recommender systems, Time series analysis, A/B testing, Hypothesis Testing, Causal analysis, Linear programming, Integer programming, Optimization