Vaibhav Sharma

Website : vaibhav47sharma.github.io

Email : vaibhav47sharma@gmail.com
Mobile : 434 - 249 - 1469

Education

- **University of Virginia** — Charlottesville, VA
  Master of Science in Data Science; GPA: 3.85 — *Jul 2018 – May 2019*
- **Birla Institute of Technology and Science, Pilani** — India
  Dual Major: MSc. Mathematics + BE. Electrical & Electonics Eng — *Aug 2010 – Jul 2015*

Professional & Technical Skills

- Python
- R
- SQL
- Java
- Tableau
- AWS
- Spark
- D3.js
- Tensorflow
- NLTK
- Scikit-Learn
- UNIX

Experience

- **Army Research Laboratory and the Data Science Institute, UVA** — Charlottesville, VA
  Data Scientist — *Sep 2018 - May 2019*
  - Developed a speech based chatbot for the cultural sensitization of Army personnel by analyzing the user cultural competence through speech input via dialogue and decision trees
  - Full-fledged product with a VR front end and a backend written in a Python Django framework
  - Leveraged NLP/text mining for feature generation - TF-IDF, LDA, word2vec, and GloVe
  - The best classification model using Gradient Boosting gave an F1-score of 0.92
  - Presented research at IEEE SIEDS 2019 conference

- **GoIbibo (One of the largest online travel e-commerce in India)** — Gurgaon, India
  Senior Software Engineer — *Apr 2018 - Jun 2018*
  - Developed a Personalized Offers Engine using models built on Logistic and SVM classifiers for customer segmentation. Led to 8% of the dropped customers returning to book a flight
  - Implemented Customer Journey analytics by developing a user tracking plan to identify customer journey touch-points. Developed 10 dashboards for six business verticals to monitor KPIs

  Software Engineer — *Jun 2015 - Mar 2018*
  - Conducted A/B testing to identify bottlenecks in the flight booking process. Implemented a wait-less booking solution that led to an increase of 1% in conversion rates, lower dropouts, improved user reviews
  - Flight grouping using sorting and hashing - used SRP (search result page) data to analyze conversion rates. Further optimized SRP results on price, baggage, booking history to increase end-user flexibility
  - Implemented MVP, Abstract Factory Pattern leading to lose coupling and increased code testability

Data Science Projects

- **Deep Learning for Detection of Neural Granger Causality in the Financial Market**
  - Leveraged neural networks (MLP and RNN) to build models that could detect Granger causality between the stock prices of 3 out of 5 firms. Models run on EMR clusters with data stored in AWS S3 buckets
- **Short-term forecasting of mid-price change in the stock market using Neural Networks**
  - Leveraged FFNN and RNN to predict change in the mid-price of a stock using NASDAQ Limit Order Book data. The best model (FFNN) resulted in 15% higher accuracy than the baseline ML models (SVM, kNN)
- **Text Mining for Machine Learning based Sentiment Analysis**
  - Built ETL pipeline to analyze and generate sentiment analysis on 0.5 million Amazon Fine Food reviews. Logistic Regression on bi-gram (Word2Vec) achieved a 93% accuracy and 0.85 AUC
- **Regression Analysis of Graduate Admissions Data to predict admission likelihood**
  - The objective of this analysis was to understand the significant factors in determining enrolment. Leveraged statistical techniques exhaustively to build a predictive regression model that could compute the probability of an admit with a mean square error of 0.31

Publications

- Paper Publication at ***2019 IEEE SIEDS Conference*** .
  Vaibhav S, Beni S, Sung Min Y, Martin B, Sodiq A, D. Brown, Erfaneh G, **Data Collection Methods for Building a Free Response Training Simulation**
  This paper details the various types of tools, techniques, and algorithms used for free speech data collection that would be used for building a free speech input chat bot