VAIBHAV SHARMA

Address: New York City | vaibhav47sharma@gmail.com | (434) 249-1469
Website | LinkedIn | GitHub

EDUCATION

University of Virginia, Charlottesville, VA Jul 2018 - May 2019

- Master of Science in Data Science, GPA: 3.85/4.00
- Relevant Coursework: Machine Learning, Data Mining, Data Visualization, Statistical Computing, Linear Models

Birla Institute of Technology & Science, Pilani, India Aug 2010 - Jul 2015

- Bachelor of Technology in Electrical & Electronics Engineering
- Master of Science in Mathematics

SKILLS

Python (pandas, Matplotlib, seaborn, scikit-learn, TensorFlow, NLTK, Gensim, Scrapy, Beautiful Soup), PySpark, SQL, SQLite, R (ggplot2, tidyr, dplyr, plotly, tm, tidytext), Java, Git, Tableau, UNIX, AWS, SAS, D3.js, A/B testing, Customer Analytics, AWS EMR, AWS S3, Spark

PROFESSIONAL EXPERIENCE

Army Research Laboratory and the Data Science Institute, UVA

Data Scientist Sep 2018 - May 2019

Published research - https://ieeexplore.ieee.org/document/8735621
Presented research at IEEE SIEDS 2019 conference

- Developed a speech based chatbot within a data-driven framework to train Army personnel to be culturally sensitive by analyzing the cultural competence of a user through speech input via dialogue and decision trees
- Developed a full-fledged product equipped with a VR front end and a backend written in Python Django framework where the speech input is converted into text, classification models are run, feedback selected, and the feedback and score sent via an API
- Leveraged NLP - Text Mining - TF-IDF, LDA, word2vec, and GloVe
- Classification Algorithms used – Logistic, Random Forest, Decision Trees, Gradient Boosting, and Neural Networks. The best model (Gradient Boosting) gave an F1-score of 0.92

Goibibo, Gurgaon, India

(Largest online travel e-commerce in India)

Senior Software Engineer Apr 2018 - Jun 2018

Analytics and Machine Learning

- Personalized offers engine
  - Developed the pipeline and the model for personalized offers after segmenting app-drop customers by using logistic regression and SVM classifier
  - Resulted in 8% of the dropped users returning to book the flight
- Customer Journey Analytics
  - Ideated and implemented a user tracking plan to identify customer journey
  - Developed 10 dashboards for six business verticals for consistent data reporting
  - Used to monitor crashes and other performance indicators
  - Resulted in the setup of user personalization engine

Software Engineer Jun 2015 - Mar 2018

Product & System Design and Android Development

- Wait-less flight booking solution
  - Conducted A/B tests to identify bottlenecks in the booking process
  - Implemented a wait-free solution on the mobile app resulting in a faster booking process
  - Led to an increase of 1% in conversion rates, lower app dropouts and improved user reviews
- Flight grouping using sorting and hashing
  - Analyzed conversion rates in the Flights vertical using search result page (SRP) product impressions data
  - Optimized SRP results on price, baggage options and booking history to increase end-user flexibility
  - Led to a 0.5% increase in click rates from SRP and a 0.3% increase in conversion rates

PROJECTS

Deep Learning for Detection of Neural Granger Causality in the Financial Market

- Leveraged MLP and RNN to detect Granger causality in stock prices of 5 firms
- Models were run on EMR clusters with data stored in AWS S3 buckets
- Models were successfully able to detect Granger causality between 3 out of the 5 stocks

Short-term forecasting of mid-price change in the stock market using Neural Networks

- Leveraged FFNN and RNN predict the change in mid-price of a stock using NASDAQ Limit Order Book data
- Models were run on EMR clusters with data stored in AWS S3 buckets
- The best model used FFNN and resulted in ~15% higher accuracy than the baseline machine learning models which used SVM and kNN

Natural Language Processing for Sentiment Analysis

- Built ETL pipeline to analyze and conduct sentiment analysis on 0.5 million Amazon Fine Food reviews
- Used Text mining techniques (TF-IDF, Word2Vec, n-grams model, LDA) for feature generation and classification models for prediction
- Logistic Regression on bi-gram (Word2Vec) achieved a 93% accuracy with 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

Android Application Development

Developed and released two android applications (Free Open Wi-Fi Connect and Wi-Fi Hotspot Tethering) on Google Play Store

URL: https://play.google.com/store/apps/details?id=com.easyway.freewifi

- The app efficiently searched and connected to free open Wi-Fi networks
- 500,000+ total downloads and a rating of 4.3/5 on Google Play Store
- Incorporated user feedback, and monitored crash and click rate to release new improved versions of the app