Aarish Irfan

aarishirfan@berkeley.edu | +1(510)-871-0377 | aarishirf.github.io/portfolio

Education

- **University of California, Berkeley** *May 2020*
  Double Major: **Computer Science** | **Data Science (Business Analytics)**
  Major GPA: ***3.5***
- **Relevant Course Work**
  Computer Science: Data Structures | Linear Algebra & Differential Equations | Discrete Math | Artificial Intelligence | Computer Security | Efficient Algorithms and Intractable Problems | Database Systems |
  Data Science: Principles of Data Science | Machine Learning | Natural Language Processing | Data Mining and Analytics | Data, Inference, and Decisions | Probability | Sports Analytics |
  Business Analytics: Analytical Decision Modelling | Engineering Economics | Economic Statistics and Econometrics |

Projects

- **Instagram Popularity Predictor (Python – Jupyter)** – a machine learning & NLP project to predict popular posts based on comments. Utilized feature engineering to devise popularity metric, SMOTE for class imbalance, Keras for embeddings and TensorFlow for LSTM model.
- **Song Genre Predictor (Python – Jupyter)** – a data analysis, machine learning & NLP project to predict song genres using lyrics. Utilized Keras for ML models (Neural Networks, Random Forests) and NLP techniques for Sentiment Analysis & TF-IDF.
- **NBA Salaries Analysis (Python - Jupyter)** – a data visualization and machine learning project based on predicting player salaries. Involved data pre-processing and feature engineering to research under/over-valued players through Regression, Neural Networks and Random Forests.
- **Tweet Analysis (Python - Jupyter)** – a data analysis project to analyze tweets of a US politician. Utilized Twitter API to gather data, Data Cleaning & Visualization through NumPy, Matplotlib & Seaborn, and NLP & VADER for sentiment analysis to research popular tweets.
- **FIFA Data Exploration (Python - Jupyter)** - a data analysis and visualization project based on the game FIFA. Involved data mining and cleaning to present stats of various players, use of graphs for visualization and linear regression modeling/ neural networks to predict player worth.
- **SFO Crime Analysis (R)** – a data analysis and visualization project based on uncovering success of crime solving. This involved analyzing peak times in criminal activity, popular crime districts and visualization/prediction of the type of crime more likely to be solved.

Work Experience

- **Data Scientist at Logos** *June 2019 - Present*
  - Developed Machine learning models and used Reinforcement Learning + NLP for an upcoming decentralized news platform
  - Worked on a comment ranking LSTM neural network based on political bias and sentiment
  - Developed a user reputation algorithm based on feature engineering
  - Used ETL processes and wrote backend scripts to migrate Firebase Database into BigQuery to enable querying for ML models
  - Wrote NoSQL queries to develop a featured article algorithm using JavaScript
  - Working on an Article Click-bait Model using NLP and Vector Spaces
- **Data Science Research Assistant at UC Berkeley Division of Data Sciences** *Summer 2019*
  - Pitched and got approval for a data analysis research project in collaboration with US Census recommending colleges from student profile
  - Use of Data Cleaning to merge large census datasets and Data Mining + Analytics to extract useful features
  - Developed a Nearest Neighbors Classification Model to predict fit colleges based on test scores, income, major, location
- **Curriculum Developer Intern at UC Berkeley Division of Data Sciences** *Aug 2018 - May 2019*
  - Jupyter Notebook Lab Development for the ESPM 163AC module using Python, Stats, Matlab and Data Science libraries
  - Discuss strategies for efficient student friendly code and extensive use of Git for their implementation
  - Draw upon prior knowledge of correlation and regression models to write efficient and student-friendly code
- **Academic Intern/Lab Assistant at UC Berkeley EECS Department** *Aug 2017 - Dec 2018*
  - Assisted TAs in organizing bi-weekly cs and data science lab sections
  - Guided students with relevant course concepts/lab work in Jupyter notebook/Python and tested their understanding
  - Suggested improvements for future semesters/addressed common problems in weekly meetings
- **Analytics Intern at Infoxcel** *Summer 2018*
  - Provided real time analytic dashboards to optimize customer experience in the financial industry by analyzing customer feedback data
  - Used open platform tools for data analysis and visualization on customer interactions through Pandas, Matplotlib, Seaborn and Excel
  - Utilized NLP and VADER to conduct sentiment analysis of the client's call records and identified any trends/patterns in data
- **Analytics Intern at TeraData** *Summer 2017*
  - Research Analysis and Visualization with an Enterprise Data Warehouse solution provider through NLP, Classification and Regression
  - Data Analysis of telecom operators using call records to determine feasible package plans for particular customers
  - Use of correlation calculation between facets of call data including client age, network, internet usage to develop marketing strategy

Skills

- **Tools & Technologies –** Scikit Learn ML | Keras | NLTK | BERT NLP | Jupyter | Pytorch | TensorFlow | NumPy | Pandas | Matplotlib | Seaborn | Excel | Git | GCP | BigQuery | Firebase |
- **Programming** – Python | Java | SQL | R | STATA |
- **Languages** – English | Urdu | French | Hindi | Punjabi |