Rohan Mahajan

1605 Rice Blvd,
Houston, TX 77005

rtm10@rice.edu
630.280.7278

EDUCATION

Rice University — **Houston, TX**
Bachelor of Science in Computer Science and Statistics — *Class of 2026*

- GPA: 3.75/4.00
- Minor: Data Science
- Relevant Coursework: Tools and Models for Data Science, Linear Regression, Introduction to Program Design, Algorithmic Thinking, Introduction to Probability and Statistics, Fundamentals of Computer Engineering

EXPERIENCE

Energy Innovation Capital — **Houston, TX**
Venture Capital Data Analytics Extern — *December 2023 – Present*

- Devised a multiple linear regression model to accurately predict start-up funding with at ~85% accuracy
- Visualized trends in 100,000+ entries of energy start-up funding data using Tableau, Python (seaborn, matplotlib), and R (ggplot2)

Valix Group LLC — **Houston, TX**
Software Engineer Intern — *October 2023 – Present*

- Currently creating a web scraper using NodeJS to gather information about commercial real estate listings in Chicago
- Engineered asset correlation software in Python for equities and futures, utilized in supporting investment decisions.

Biokind Analytics — **Houston, TX**
Data Consultant — *September 2023 – January 2024*

- Partnered with Halo House, a healthcare non-profit focused on offering housing to blood cancer patients
- Analyzed data and found trends in 30,000+ rows of patient and donation data contextualized with seasonality trends to visualize (GeoPy, seaborn, ggplot2) and identify critical periods of funding, and proposed deployment methods

PROJECTS

Fake News Detection using Natural Language Processing GitHub — **Westmont, IL**
Python — *August 2023*

- Preprocessed 6,000+ texts by eliminating stop words, tokenizing, and padding data
- Utilized word embeddings from GloVe text file to generate word embeddings matrix specified for texts
- Engineered multi-layered neural network using TensorFlow and predicted validity of news at ~90% accuracy rate
- Employed Term Frequency-Inverse Document Frequency (TF-IDF) Vectorizer, creating relative word frequency matrix
- Applied Logistic Regression to classify data from TF-IDF vectorizer, predicted validity of news with > 92% accuracy

Music Recommendation using Unsupervised Learning GitHub — **Westmont, IL**
Python — *July 2023*

- Conducted extensive data analysis and visualization on 100,000+ songs along with their features
- Utilized silhouette, elbow methods to identify optimal cluster count, employed K-means to cluster both song and genre
- Implemented Principal Component Analysis (PCA) to reduce dimensionality in order to visualize clusters

Discounted Cash Flow Automation with Monte Carlo Simulation — **Westmont, IL**
Python — *March 2023*

- Data scraped (requests) to find values corresponding to be utilized in Discounted Cash Flow valuation
- Ran 20,000+ simulations based on normal probability distribution for variable inputs for model
- Modeled valuations through a plot which illustrated the inputted company's valuation in each simulation and the relative frequency of each valuation

Network Resilience Analysis — **Houston, TX**
Python — *February 2023*

- Assessed network resilience of Erdos-Renyi and Upa graph models as a graph representation of a real world network
- Implemented algorithms to simulate both random attacks and targeted attacks on highly connected nodes to observe the robustness of different topologies in real world environments such as transportation or supplier networks

MLB Win Prediction Regression — **Houston, TX**
Python — *October 2022*

- Created linear regression model using Least Absolute Shrinkage and Selection Operator to predict future performance of an MLB team utilizing 122 aggregate team statistics collected from 1954 - 2000
- Evaluated model performance using Mean Squared Error and visualized results using matplotlib

TECHNICAL SKILLS

- **Languages:** Python, Java, SQL, R, HTML, CSS, NodeJS
- **Technologies:** pandas, scikit-learn, TensorFlow, NumPy, matplotlib, NLTK, seaborn, plotly, GeoPy, ggplot2
- **Tools:** Jupyter, Virtual Studio Code, IntelliJ, GitHub, Tableau, Excel