ARNAV SHAH

33 S 3rd St | San Jose, CA 95113 | 845-675-2881 | arnav.shah@nyu.edu
https://www.linkedin.com/in/arnavshah96 | https://github.com/arnav1996

EDUCATION

New York University, Tandon School of Engineering, New York, NY
Master of Science, Computer Science, GPA: 3.22 May 2019
Relevant Coursework - Design and Analysis of Algorithms, Foundations of Data Science, Big Data, Computer Vision and Scene Analysis, Machine Learning, Cloud Computing.

Gujarat Technological University, LD College of Engineering, Gujarat, India
Bachelor of Engineering, Information Technology, GPA: 3.60 June 2017
Relevant Coursework - *Artificial Intelligence, Data Mining and Business Intelligence, Distributed Operating Systems, System programming, Computer Networks, Operating Systems, Database Management Systems.*

TECHNICAL SKILLS

Languages: Java, R, C, C++, SQL, Python, HTML5, CSS3, PHP, JavaScript
Tools: Hadoop, MapReduce, Hive, Pig, Spark, AWS S3, AWS Lambda, Kafka, Tableau, React, Git, Github
Other: Linux/ Unix, Windows 10, MongoDB, MySQL, NoSQL, Oracle, Data Structures

EXPERIENCE

Intern, AllocateRite LLC, New York, NY,USA May 2018-Present

- Wrote Python scripts which can run periodically to generate tear sheets which lists out the returns and various Statistics by using Reportlab and Matplotlib which can fetch data (data for tear sheets) from the MySQL database which is present on the EC2 instance of the AWS to generate graphs using Matplotlib. It enabled generation of tear sheets in PDF format after rebalancing.
- Analyzed the website and dashboard of the AllocateRite and suggested ways to improve the layout and graphics of the website to make it interactive.
- Building a website using React, JavaScript, HTML, Bootstrap and CSS for the payment platform.

PROJECTS

Temporal and Weather Data Analysis on Citi Bike Trips May 2018

- Explored the variance in ridership due to temporal factors and weather patterns like temperature and precipitation. Identified the most significant relationships between two datasets using spearman's correlation coefficient.

Exploring 311 Service Request Data [Jupyter Notebook, Python] December 2017

- Performed predictive analysis for resolution time and time series analysis for forecasting the volume of complaints for 311 phone service.
- Calculated the yearly, weekly and daily seasonality of the time series data using Prophet model. Used Random Forest and Decision Tree classifiers for classification. Used appropriate evaluation metric for each model.

Credit Card Fraud Detection System [Jupyter Notebook ,Python] May 2017

- Performed various steps such as exploring data, building neural network and visualizing data. Used Tensor Flow to build the neural network and used t-SNE to visualize the data. Achieved accuracy of 83% in detecting fraudulent transactions.

Twitter Sentiment Analysis and Word Count [CDH] May 2017

- Performed Real time fetching of sentiments using appropriate keywords from Twitter using various tools such as Flume, HDFS, Hive and Hue. Created a new table to include the sentiment value of each tweet.
- Performed word count in a text file using Cloudera and displayed frequency of words present in file.

STUDENT ACTIVITIES

***Volunteer*, Cybersecurity Awareness Week 2017**, New York University

- Assisted student attendees in checking-in at the event.
- Provided information on event agenda and answered questions from guests

***Participant*, AMA** (Personality Development and Effective Speaking workshop)

- Improved communication skills and learn about topics related to personality

***Key Player*, Cricket Tournament, Rajpath Club**