**Avneet Singh**

as696@njit.edu | (732)-996-9632 | Freehold, New Jersey | LinkedIn: [in/AvneetSinghURL/](https://www.linkedin.com/in/avneetsinghurl/) | Kaggle: Asingh278

**Objective:** Seeking entry level Data Analyst or Data Scientist position to gain industry experience and expand skillset

**Education**

**New Jersey Institute of Technology** – *M.S. in Data Science*  **Expected May 2023**

**Coursework:** Big Data, Data Analytics with R, Data Mining, Applied Statistics *GPA: 4.0*

**Activities:** Google Developer Student Club (Member)

**Stevens Institute of Technology** *– B.S. in Quantitative Finance*  **Graduated May 2020**

**Relevant Coursework:** Data Management in R, Applied Models and Simulations, Discrete Structures, Data Structures, Machine Learning, Algorithms, Probability and Statistics

**Skills**

**Programming & Databases**

* Python (pandas, NumPy, MRjob, Spark MLib)
* R (dplyr, ggplot2, plotly)
* Spark
* MySQL
* NoSQL (HBase, MongoDB)
* Hadoop Distributed File System (YARN, M/R)

**Machine Learning Models**

* Classification
* Regression
* Support Vector Machines
* Feature Selection
* Decision Trees
  + Random Forest
* Natural Language Processing
* Data Clustering
  + kNN, K-Means

**Software Applications**

* Tableau
* Microsoft Excel
* Jupyter Notebook
* AWS (EC2)

**Soft Skills**

* Attention to detail
* Problem Solving
* Effective Oral and Written Communication

**Project Experience**

**Predict Future Sales Kaggle Competition (R)** *December 2021*

* Forecasted next month’s total sales for 214,000-row test dataset via SLR and MLR machine learning models
* Discerned 2 optimal features for predictive linear regression model via sequential forward selection (SFS)

**Twitter Sentiment Analysis via Spark Streaming (Python)** *December 2021*

* Utilized Twitter API to fetch relevant tweets and classify sentiment as positive (4) or negative (0) via Spark MLib
* Trained Logistic Regression Machine Learning model to perform sentiment analysis on 1,600,000-row dataset
* Stored fetched tweets’ data and classification results to external database and yielded 82.6% ROC accuracy

**Equity Portfolio Management (R)** *October 2021*

* Calculated mark-to-market values for 5-day Buy Low & 5-day Buy High investment strategies for 5 tech stocks
* Compared MTMs to High-Tech Index, converted stocks from USD to JPY, & generated updated MTM values
* Generated optimal strategy that maximizes MTM (Buying Low, interval = 21 days, MTM = $5,771,743)

**K-Means MapReduce Development and Execution via AWS EC2 Instance (Python)** *September 2021*

* Initialized and configured Hadoop environment via AWS with Java, Spark, and Python installations
* Generated 9,000 (X,Y) coordinates based on 3 predetermined initial centroids
* Developed and executed a K-means MapReduce class via MRJob to discover centroids of clusters

**Predictive Modeling for FDA Approval of Phase III Drug-Indication Pairs (Python)** *May 2020*

* Accumulated and polished comprehensive 606x9 dataset to quantify and analyze key data features, including approval percentages, average trial duration, category of drug application, etc. in support of research project
* Executed statistical analyses via utilizing Binary Classifiers (Gaussian, Zero-R) and ML-supervised models (5-Fold 3NN, Random Forest, Logistic Regression) to test hypothesis, determine correlation, and develop solutions

**Work Experience**

**Stevens Institute of Technology |** Hoboken, NJ *September 2019 – May 2020*

*Office Assistant*

* Analyzed public data and compiled statistics on 2,000 qualified tennis players to determine potential recruits
* Cleaned and maintained data from public tennis ranking databases in support of recruitment efforts