APOORVA JAIN

LinkedIn | 469-623-5970 | apoorvaj@smu.edu | Github

SUMMARY: Looking to apply hands on experience utilizing predictive analytics and classical modeling techniques for statistical inference and create insights. Expert communicator with proven ability of working with cross-functional teams.

EDUCATION:

Southern Methodist University, Dallas, Texas **Aug 2020**
Masters in Engineering Management specializing in Information System GPA: 3.1
Masters in Network Engineering GPA: 4.0
Rajiv Gandhi Prodyogiki Vishwavidyalaya, Bhopal, India
Bachelors in Electrical and Electronics Engineering **June 2017**

SKILLS:

Competency Areas: R, Python, SQL, Descriptive Analytics, Predictive Analytics, Inferential Analytics, Exploratory Analytics, Content Analytics, Neural Networks (CNN, RNN)
ML Algorithms: Cluster Analysis, Regression Analysis, Decision Tree, Random Forest, Naïve-bayes, Supported Vector Method, Time Series, A/B Testing, ETL
Data Science Tools: RStudio, Jupyter, Talend, Tensor Flow, Minitab, SPSS, Advanced MS Excel , MS Power BI, Tableau
R : dplyr, ggplot2, caret, ROCR, lubridate, lattice, CaTools, text mining, tidy verse, mlr, e1071
Python : NumPy, Pandas, Scikit Learn, Seaborn, Matplotlib, Keras
Databases: MySQL, Redshift, Microsoft SQL Server, Hive

PROJECTS:

COVID 19: Predictive Analysis(Power BI, Python) **Feb. 20 – Apr. 20**
- Validated 1.8 million datasets and predicted by supported vector method
- Created data visualizations such as geographical, tree and calendar map with confirmed cases, deaths and recovered cases of COVID-19 epidemic on Python

Netflix Movie: Recommender Engine(Python, R) **Jan. 20 – Mar. 20**
- Preprocessed 7.5 million data to develop content-based movie recommender system
- Optimized the number of clusters to reduce the heterogeneity till 15% using K-means method
- Suggested movies based on what the user liked and the rating of the movie and cross validated - RMSE 0.8677

Sentiment Analysis: Content analysis(R, SQL) **Dec. 19 – Jan. 20**
- Developed different machine learning model on R to perform sentiment analysis and data categorization over the tweets of GOP debate that held in Ohio from SQLite database
- Evaluated the model performance using confusion matrix, ROC and AUC curve with an accuracy of 98.67% and 68.3%, determined and visualized the frequency of candidate that was mentioned the most and the sentiment of the tweet

Ilumexico: Model for Solar Plant Company (R, TABLEAU) **Apr. 19 – May. 19**
- Performed data exploration using R, adjusted for missing values and anomalies, identified outliers
- Build machine learning models using Cluster Analysis (k-means:14) and Decision Tree to enhance payment portfolio of the product and determined good indicator for customer compliance to payment schedule - "internal credit rating"
- Predicted most important features in determination of the classification variable, using Logistic Regression and evaluated the model performance using confusion matrix, ROC and AUC curve with an accuracy of 72%.

EXPERIENCE:

Altshuler Learning Enhancement Centre (SMU), *Tutor*, Dallas, USA **Sep. 19 - Current**
- Provide academic assistance to the students for statistics and probability field to achieve better understanding practically and conceptually using different statistical tools
- Encourage critical thinking, develop communication and leadership skills

Disability and Success Strategies (SMU), *Student Assistant*, Dallas, USA **Aug. 18 - Current**
- Created 100's of organized student reports for the administrative purpose using MS Excel and Box
- Handled the administrative work from enrolling students to providing solutions to their issue

Vidushi Infotech, *Analyst*, Pune, India **Aug. 16 – Jul. 17**
- Performed data wrangling and data aggregation using R, checked for missing values and identified anomalies in the data
- Modelled and automated reports of operations and marketing team to reduce the costs and increase the scalability
- Created 10 plus KPI dashboards on Power BI to report the cost-effective solutions