Arpit Chaukiyal

M (469) 929-4740 | chaukiyal.arpit@gmail.com | https://www.linkedin.com/in/arpitchaukiyal/ | https://github.com/arpitchaukiyal/

EDUCATION:

Master of Science, Business Analytics
The University of Texas at Dallas
Aug '17-Date | **GPA: 3.89/4.00**

Related Coursework:

- Business Analytics
- Advance Business Analytics
- Applied Econometrics & Timeseries Analysis
- Statistics and Data Analysis
- Big Data Analytics
- Business Data Warehousing

Bachelor of Technology, Information Technology
Uttar Pradesh Technical University
Aug '08- Jun '12

Related Coursework:

- Advanced Mathematics I & II
- Computer programming

SKILLS:

Analytical Programming Skills:
R, Python, R-shiny, Java.

Software:

Tableau, Advance Excel, R-Studio, Eclipse, MS Office, Python Notebook, PyCharm, Git.

Statistical Techniques:

Linear Regression Analysis, Logistic Regression Analysis, T-testing, ANOVA, Social Network Analysis, Market Basket Analysis.

Machine Learning Algorithms:
SVM, Decision Trees, KNN, Principal Component Analysis, Clustering (K-means), Naïve Bayes, Natural Language Processing.

Big Data Ecosystem:

Hadoop, HDFS, Flume, Hive, Pig Latin, Spark, Linux

CERTIFICATIONS:

Completed.

- Big Data and Hadoop. **Edu Pristine**
- **Data Science Associate**. **EMC²**
- Data Science with 'Python'. **UTDallas**
- R Programming. **J. Hopkins University**
-

On Going.

- AWS – SAA **AWS**
- Spark & Hadoop Developer. **Cloudera**

NOTABLE PROJECTS:

Natural Language Processing

- **ASK:** Developed a logistic regression model that **predicts the rating**. | Given: URL of cars.com.
- [R and R Shiny]: Developed a shiny application that calculates Sentimental Analysis and **predicts rating** for Toyota Camry cars **with 78% accuracy**. Sentiment scores are calculated from user reviews scraped from www.cars.com website pragmatically.

Sentiment Analysis of Twitter Feeds

- **ASK:** Calculated AFFIN scores of tweets parsed from twitter JSON file. | Given: Twitter account details.
- [Flume, HIVE, Hadoop]: Configured **Flume for stream the twitter** data into HDFS. Flatten the twitter JSON data into **Hive table using serde-jars** and calculated the AFINN score for analysing the sentiments of all tweets and re-tweets.

Social Network Analysis

- **ASK:** Shiny app created to study social network of emails shared within European union. | Given: Email data from a large European research institution and their corresponding departments.
- [R, R Shiny]: Analysed and visualized the network data of email exchanges from a European research institution having more than 1000 nodes and 25000 vertices. Calculated matrices like **2-hop neighbours, degree &, 'betweenness' centrality**, to get the insights.

Security classifier

- **ASK:** Visualization and generated insights using tableau from the GT database. | Given: Database containing all attacks in 40 years.
- [R, Tableau]: Created visualization and generated insights using tableau from the GT database. Used unsupervised learning techniques like **k-means** and mahalanobis to create clusters of countries that are vulnerable. Created a **logistic regression model** to predict which countries are more prone to attacks with 81% accuracy.

Taxi Choice Report

- **ASK:** Analysed and developed predictive model for taxi choice.| Given: New York city Taxi and Limousine Commission historical data.
- [R, MS- Excel]: Analysed the New York city Taxi and Limousine Commission historical data using R to find answers like the factors that affect the frequency of cab services, the choice of a particular taxi type by customers.

WORK EXPERIENCE:

Actuarial Analyst in **Mphasis** **Jun '16 – Jul '17**

- Developed data solutions based on **predictive and behavioural models** using regression and classification techniques like **decision trees & nearest neighbour** for life insurance client which optimized their business inflow by 17%.
- Developed model that predicts the insurer's **fraudulent behavior** in paying the premium as well as estimating the premium amount depending on undertaking data resulting in 11% reduction in defaulter volume.

Data Analyst in **Birlasoft** **Nov '14 – Jun '16**

- Liaised with the client team to develop a **model that predict the work-hours** required to complete the task depending upon resources involved and the complexity of the task to align the project with CMMI-5 standards resulting in a 15% increase in team's efficiency.
- Incorporated **linear, logistic regression & random forest techniques** to develop statistical models for demand forecasting and procurement analytics which had in decreased the procurement cost by 7% annually.
- **Trained 30 plus** employees over the period of 8 months in technical and process-oriented modules which reduced on-boarding time by up to 20%.

Software Developer in Birlasoft **Oct '13 – Nov '14**

- Developed various in-house utility modules & modified the existing processes to increase productivity (including requirements analysis, coding, unit/functional testing, deployment of test staging environments, for UAT).