SHAMA ZABEEN SHAIK

(803)-477-5794 | skshama18@gmail.com | www.linkedin.com/in/shama-zabeen-shaik | Charlotte, NC 28262

- A data science enthusiast with 3.5+ years of academic and over 2 years of industrial experience in data analytics, machine learning and big data technologies.
- Solid understanding of statistical models, machine learning algorithms and mathematical concepts with wide range of experience in Data Collection, Preparation, Exploration, Model Building, Evaluation, Data Visualization and Statistical Modeling.

EDUCATION

Master of Science in Computer Science | University of North Carolina Charlotte, Charlotte NC **GPA: 3.9/4.0**
Bachelor of Technology in Computer Science and Engineering | VIT University, Vellore, TN, India **GPA: 4.0/4.0**

TECHNICAL SKILLS

- **Programming Languages:** R, Python, SQL, HTML, CSS, JavaScript, C++
- **Tools**: RStudio, Python, Anaconda, Jupyter Notebook, Tableau, Dash by Plotly, Knime, Lisp Miner, Orange, Qlikview, XL Miner
- **Machine Learning:** Classification, Linear Regression, Logistic Regression, ROC, Decision Trees, Random Forests, Neural Networks, SVM, Naïve Bayes, K-Nearest Neighbors, K-Means Clustering, Hierarchical Clustering, Deep Learning, Text Analytics, Twitter Data Analysis, Sentiment Analysis, Time Series Data Analysis
- **Microsoft Office –** Excel (lookups, Pivot Tables, Macros, VBA), Word, Access, Visio
- **EPR system and Reporting:** SAP, Oracle, JDA

EXPERIENCE

Data Analyst **Continental Tire, Fort Mill, South Carolina** **Dec 2018 - Present**

(Python, R, Tableau, Time Series Analysis, Random Forest, Regression, Clustering, Neural Networks)

HR Predictive Modelling

- Having seen a lot of employees leave the company and not knowing the reason behind increasing attrition rate in each department, I started to analyze the company's attrition rate in each department and location of continental.
- As a part of this project, I dealt with fetching the Continental employee data from the server to cleaning, preprocessing and analyzing the insights of the attrition rate.
- I created dashboard views of the employee attrition rate and also predicted the likelihood of the employee attrition
- Since hiring involves a lot of cost, my model helped in reducing the hiring cost by proposing the ideas to retain employees with the company.

Gender Diversity Analysis

- Analyzed the data for gender pay ratio, employee count ratio by gender, gender diversity by age group, by job title, salary group and job location, very interesting insights were drawn and were presented in the form of an interactive dashboard.
- It was observed that there was a significant difference between the male and female employees in every field that was analyzed.
- After presenting the results and walking the board of senior executives through the analysis, improvements to the business plan were made that has got down the diversity rate by a very significant number.

Tele Commute (Work from Home) Analysis

- Continental introduced tele commute (work from home) facility at the beginning of 2018. All the managers and the executives were very curious on finding out the tele commute utility ratio and the trends in the work from home days utilized.
- After a significant amount of analysis, results like which department is using the telecommute to the fullest, utility ratio of days, utility ratio of employees by gender, by sub group, by age group, by job location, by cost center, year to date utility rate were drawn and were hosted on the Continental server in the form of an interactive dashboard.

Graduate Research Assistant **University of North Carolina, Charlotte, North Carolina** **Dec 2017 – Dec 2018**

(Python, R, HTML, CSS, PHP, Time Series Analysis, Classification, Forecasting)

- Trend and competitive Analysis on Continental's Market and Pricing data for the Americas.
- Forecasting the future trends in the product prices by bench marking the desired brand and features.

Database Analyst **Tecra Systems Pvt Ltd, Hyderabad, India** **Dec 2016 – July 2017**

(HTML, CSS, PHP, JavaScript, MySQL, Data Base)

- Data base schema design, data collection, data base creation and web designing for Sri Vivekananda Public Schools (India)
- Created a database for all the student and faculty records available.
- Created a website with different login pages for faculty, staff and students for the school using JavaScript, MySQL, HTML, CSS.

ACADEMIC PROJECT WORKS

Sentiment Analysis on Political Twitter Data *(R, Tableau, Sentiment Analysis, Data Pre-processing, Text Analysis, Topic Modelling, LDA)*

- The main aim of the project was to find variation of sentiment between democrats and republicans of the US House of Representatives with respect to the trending topics.
- Performed Data preprocessing on the twitter political data from year 2009-2010 using Microsoft Excel and Tableau.
- Segregated the data into Republicans and Democrats to perform sentiment analysis. Performed Topic modelling using LDA Tuning and LDA to find out the topics in the datasets.
- Applied Lexicon based sentiment analysis using 'SentimentR' package and 'Syuzhet' to classify the sentiment into positive, neutral and negative of both Republicans and Democrats with respect to discovered topics.

New York City Parking Ticket Analysis **Jan 2018 – Apr 2018**

(Python, XL Miner, Naïve Bayes Classification, Clustering, Association Rule Mining, Dimension Reduction, Data Normalization)

- Our hypothesis of the project is to find out where and when, a greater number of violations are occurring for a given county, vehicle type and time in the New York City.
- Pre-processed huge dataset by applying dimension reduction and normalizing the data using Microsoft Excel.
- Incorporated different prediction, classification, clustering and association rule mining models using XL Miner.
- Applied Naïve Bayes classification algorithm on the dataset to get approximate results for the hypothesis considered.
- Predicted with 76% accuracy that Kings County (Brooklyn) is the area most prone to violations especially in the months of June, October and September. The model also suggests that Friday's are the worst days where most traffic violations occur.

HR Predictive Modelling **Aug 2017 – Dec 2017**

(Python, R, Dash by Plotly, Decision Tree, Random Forest, SVM, ADA Boost, Linear regression)

- Applied Predictive Data Analytics on Attrition, Absenteeism and Time to hire data from Continental AG. Implemented Data cleaning using R Programming Language and Microsoft Excel.
- Performed exploratory data analysis with the help of Tableau data visualizations.
- Developed Logistic Regression, Random Forest for the data analysis to build models and predictions.
- Random Forest Algorithm applied for prediction of hypothesis was 75% accurate.
- Created interactive dashboard with python framework Dash by Plotly which was coded and deployed on Heroku for external access.

Co-Creative Robot **Jan 2018 – May 2018**

(R, Python, NLP, Tensor Flow, Clustering, Named Entity Recognition, Speech to Text Conversion, CNN, Text Analysis, Sentiment Analysis)

- Trained the robot to sketch the scenario on canvas based on the voice inputs of the user and further developed the program by updating the image with user feedbacks.
- Used Spacy in R and Python NLTK for grouping words and creating clusters from the voice inputs.
- Deployed sentiment analysis for vector representations of text and response classification to develop the feedback module.

Black Jack game using Reinforcement learning **Apr 2018 – May 2018**

(Python, Reinforcement Learning, Tensor Flow, Classification, Machine Learning)

- I have made use of the Q-learning model to predict the action that could be taken for a given state based on the previous learning with the help of rewards.
- Blackjack is a card game commonly known as twenty-one. It is played between multiple players and a dealer. Each player competes against the dealer but not against each other. The objective is to beat the dealer in one of the following ways:
- Get 21 points on the player's first two cards (called a "blackjack" or "natural"), without a dealer blackjack.
- Reach a final score higher than the dealer without exceeding 21.
- Let the dealer draw additional cards until their hand exceeds 21.

Rapidly Growing Random Tree Planner **Apr 2018 – May 2018**

(Python, Machine Learning, Rapidly Growing Random Trees, Motion Planning Algorithm)

- Rapidly Growling Random Trees, one of the path planning algorithms is made use of in this project.
- An RRT is an algorithm designed to efficiently search non-convex, high-dimensional spaces by randomly building a space-filling tree which is constructed incrementally from samples drawn randomly from search space and are biased to grow towards the large un searched areas of the configuration space.
- Implemented the project in python and tried to reduce the computation speed of the path from the start state to the goal state.

On-Campus Involvement

- 2nd in the Innovative Computing project showcase for the HR Predictive Modelling Project **Dec 2017**
- Awarded the best project among 6000 for Weather Data Analysis Project. **May 2017**
- Poster submission for IEOM-2016 (Malaysia International Conference) **Dec 2016**