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 1.5 years of industrial experience in data analytics, machine learning, and big data technologies.
- Proficient knowledge in statistics, machine learning algorithms, and analytics with excellent understanding of business operations and analytics tools for effective data analysis.
- Extensive hands on experience and high proficiency with structured, semi-structured and unstructured data, using a broad range of data science programming languages and big data tools including R, Python, and SQL.
- Expertise in transforming business requirements into analytical models, designing algorithms, building models, developing data mining, and reporting solutions that scales across massive volumes of structured and unstructured data.

EDUCATION

Master of Science in Computer Science | University of North Carolina Charlotte at Charlotte, NC **(GPA: 3.9/4.0)** **Dec 2018**

Bachelor of Technology in Computer Science and Engineering | VIT University, India **(GPA: 4.0/4.0)** **May 2017**

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, 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

EXPERIENCE

Data Analyst | Continental Tire, Fort Mill, SC **May 2017 - Present**

- **HR Predictive Modelling**
  - Fetch employee data from the server to clean, pre-process, and analyze the insights of the attrition rate. Predict the likelihood of the employee attrition rate and create dashboard views of analysis and predictions.
    - ✓ **Outcome:** Suggested a new tactic to persuade leaving employees to stay with the company, resulting in a 5% decrease in attrition.
- **Gender Diversity Analysis**
  - ETL operations on data using complex SQL queries. Interpret & analyze the results using statistical techniques like MS EXCEL, Tableau. Provide ongoing reports and display in front-end dashboards created using Tableau, R-shiny, Python, MySQL.
    - ✓ **Outcome:** Introduced a live connection to the dashboard with auto fetching of data from the database, resulting in reduced manual data fetch and load operations.
- **Tele Commute (Work from Home) Analysis**
  - Scripted a priority scheduling algorithm in Python to prioritize the work from home eligibility. Analyzing the tele commute utility ratio and the trends in the work from home days utilized.
    - ✓ **Outcome:** After introducing the priority scheduling, the tele commute rate increased slightly by 0.8% with in a month time frame.

Graduate Research Assistant | University of North Carolina at Charlotte, NC **Dec 2017 – May 2018**

- Scrapped data from web using Python, designed database schema and created NO SQL database. Introduced dynamic interactive dashboards that replaced traditional static reports, increasing the productivity of sales and helping in better understanding of the data insights.
  - ✓ **Outcome:** Trend and competitive analysis on Continental's Market and Pricing data for which resulted in an increase of 16% sales of the winter tires.

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

- Data base schema design, data collection, data base creation and web designing using JavaScript, MySQL, HTML, CSS
  - ✓ **Outcome:** Created a database using MySQL and saved the information of about 10,000 students and 280 staff with different logins for both.

ACADEMIC PROJECT WORKS

Sentiment Analysis on Political Twitter

- Performed Data preprocessing on the twitter political data from year 2009-2010 using Microsoft Excel, Tableau, R and 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 on the topics discovered from LDA.

New York City Parking Ticket Analysis

- Pre-processed huge dataset by applying dimensionality reduction and normalizing the data using Python and XL Miner.
- Tested prediction, classification, clustering, and association rule mining models using XL Miner.
- Applied Naïve Bayes classification algorithm on the dataset to get results for the hypothesis considered with an accuracy of 76%.

HR Predictive Modelling

- Applied Predictive Data Analytics on attrition, absenteeism, and time to hire data from the Continental AG. Implemented data cleaning using R Programming Language and MS Excel.
- Random Forest Algorithm applied for prediction of hypothesis which was 76% accurate.
- Created interactive dashboard with Python Framework Dash by Plotly coded and deployed on Heroku for external access.

Co-Creative Robot

- 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.