[**SHAMA ZABEEN SHAIK**](https://skshama18.wixsite.com/mysite)

**+1 (803) 477-5794 |** [**skshama18@gmail.com**](mailto:skshama18@gmail.com)**|** [**linkedin.com/in/shama-zabeen-shaik**](http://www.linkedin.com/in/shama-zabeen-shaik)**|**[**shamask.github.io/**](https://shamask.github.io/) **| Charlotte, NC**

* A Data Science enthusiast with over **4+ years** of **academic** and **3+ years** of **industrial** experience in **Data Analytics**, **Machine Learning,** **Text Analytics, Predictive Analytics, NLP** & **Building Statistical models**.
* Strong **Communication** and **Analytic skills** with the ability to extract, collect, organize, analyze and interpret trends or patterns in complex data sets.
* Proficient knowledge in **Mathematics** and **Statistics** with excellent understanding of **Sentiment Analysis, Survival Analysis, Donor Analytics, Predictive Analytics** and **Text Analytics.**
* Extensive hands-on experience in using a broad range of data science programming languages and bigdata tools including **Python, R, Tableau, SQL, Excel, ERP.**
* Working knowledge of **Agile, Scrum Methodology** and **Waterfall Model** in project development.
* Basic Knowledge on Amazon’s services, **EMR, Redshift, S3** and **EC2**
* Three time Data Science Hackathon **Winner** Feb 2020, Feb 2019 and Mar 2019, organized by **Red Ventures** and **American Tire Distributors**

**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
* **Tools**: RStudio, Python, Anaconda, Jupyter Notebook, Tableau, Dash by Plotly, Knime, Lisp Miner, Orange, XL Miner, Weka, MySQL, PostgreSQL, GitHub, Microsoft Excel, Microsoft Visio, Splunk
* **Machine Learning:** Linear Regression, Classification, Logistic Regression, Decision Trees, Random Forests, SVM, Hypothesis Testing, K-NN, K-Means Clustering, Text Analysis, Sentiment Analysis, Time Series Analysis, Survival Analysis, A/B Testing, Numpy, Matplotlib, Sckit-learn, Pandas, Scipy, scikit-learn, caret, mlr, mllib
* **NLP techniques: T**okenization, Part-of-speech Tagging, Parsing, Stemming, Lemmatization, Semantic Analysis, Named Entity Recognition, Modeling and Work Representations, RNN, ConvNets, TF-IDF, LDA, Word2Vec

**EXPERIENCE**

**Data Analyst |** Wells Fargo, Charlotte, NC **Mar 2019 – Present**

* Create Daily, Weekly, Monthly, Quarterly and Annually illustrative interactive dashboards to monitor Platform health using Splunk and Tableau. Analyze the incidents by the root cause and predict the potential failures on a selected application.
  + **Outcome:** The model helped enhance the resource utilization by 50% and helped in cutting down the maintenance time from 12 hours to 4 hours every weekend.
* Create live interactive single platform view dashboards to monitor emergency fixes, status updates for critical processes and job runs, abends by the root cause, ad-hoc requests to be addressed.
  + **Outcome:** Rather than using 5 different applications to monitor a job or an application, this dashboard platform gives user the privilege to view all of the requirements in a single page, Thus saving tremendous time and manual efforts.
* Design Database schema to store the SLA compliance metrics for any given job/application at a specified time of the day. Create interactive dashboards to monitor SLA compliance.
  + **Outcome:** This dashboard gives user a clear picture of the jobs that are about to miss an SLA and notifies the user through mail when the time for SLA compliance is <25% remaining.
* Work with NLP libraries such as NLTK, openNLP, Stanford-NLP, WordNet, SAS Text Miner or NLP software to track system progress and improve upon existing methodologies by developing new data sources, testing model enhancements, and fine-tuning model parameters.

**Data Science/ Machine Learning Student Intern|** Continental Tire, Fort Mill, SC **May 2018 – Dec 2018**

* **HR Predictive Modelling**
* Work with Business Leaders to gather the KPI’s, fetched employee data from the server to clean, pre-process, and analyze the insights of the attrition rate. Predict the likelihood of employee attrition rate, create interactive 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**
* Using complex SQL queries, performed ETL operations on data. Interpreted & analyzed 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% wit 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 & competitive analysis on Continental’s Market Pricing data resulting an increase of 16% sales of winter tires.

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

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

**ACADEMIC PROJECT WORKS**

**Sentiment Analysis on Political Twitter** | University of North Carolina at Charlotte, NC **Aug 2018 – Nov 2018**

* 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.

[**HR Predictive Modelling**](http://my-dash-hrpred2.herokuapp.com/) | University of North Carolina at Charlotte, NC **Aug 2017 – Nov 2017**

* 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**](https://skshama18.wixsite.com/website-1) | University of North Carolina at Charlotte, NC **Jan 2018 – Nov 2018**

* 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.

**ACHIEVEMENTS**

**Queen City Hackathon Feb 2020**

## First place winners of Charlotte's largest data science and machine learning Queen City Hackathon among 300 participants by providing potential solution for the socio-economic data by building a Machine learning model Light Gradient Boost Regressor.

## Utilized Amazon Sage Maker and created a Jupyter Notebook to create the Machine learning model Light Gradient Boost Regressor which processed the data from S3. The Sage Maker model was used to compare and improve the efficiency of the model created.

**Reinvent the Wheel - ATD Hackathon Mar 2019**

## Bagged the First Prize in Predicted demand for replacement tires across the ATDs Distribution Centers, by using a Machine Learning algorithm Light Gradient Boost and presenting it graphically by creating Tableau Dashboards.

**Queen City Hackathon Feb 2019**

## First place winners of Charlotte's largest data science and machine learning Queen City Hackathon among 300 participants by providing potential solution for the social cause Opioid Crisis (Healthcare Analytics) by building a Machine learning model Light Gradient Boost classifier. Demonstrated a data story on the health care analytics data using the Business Intelligence Tool Tableau.