**MOHD MOHI UDDIN** 

**Ph : 9398635877 Email:****mohd.mohiuddin.549@gmail.com**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Osmania University, Hyderabad, India** MAY2013– MAY-2016

Bachelor of Commerce (B.com)

**Google Cloud Proficiency Badges:**

* Proficiency Badges from Google cloud for Engineer Data in Google Cloud
* Proficiency Badges from Google cloud for Insights from data with Big Query

**Udemy courses:**

* Certification in Natural Language Processing with python
* Certification in Linux Command Line Basics

**Altreyx Core Designer :** Certified Alteryx Designer Core

**Nation Stock Exchange**

Certification from NSE in Business Analytics

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

 **Skills and Courseware** **Languages & Tools :** Python, PLX Google SQL,App Script,Tableau,Alteryx,Data Studio,Excel, Google sheets,Bigquery, Data loss prevention tool and GCS. Python **Libraries:** Pandas, Numpy, Scikit-Learn, Spacy, Ploty, Keras and Tensorflow. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ **Work Experience**

**Data Analyst, Global Logic , Hyderabad (June-2019-Present)**

Part of the Central Analytics Team for internal Project areas of Research data platform operational services. Working as a data analyst using Tools such as SQL, Python colab, Data studio and Google Sheets,Shell scripts and google cloud services( Big query and custom cloud function for plugin).

* **Autonomous cars Labeled Data Dashboard :: Plx (Google Sql) and Data Studio :**

The aim of this dashboard was to displays the efforts and labels generated by labeling resources for all related data sources In 4 fields (i.e. Image, Audio,Video, Annotation,Lidar).

Impact : Helped Pgm to understand and present the data worked by human raters to the team.

* **WorkForce Management Dashboard :: Google Sql and Data Studio:** Created a dashboard that would help to keep track of tasks and forecast the completion of tasks which will help in better optimization of resources and maintain our trust with our clients and welcome new clients.

Impact : Helps TL and managers to keep track of human raters performance and optimizing tasks better.

* **Executive Dashboard :: Google Sql and Data Studio:** Created a dashboard that would help to update the stakeholders with the status of volume completion and forecast with completion dates for different Project Areas. This dashboard operates 24 Project areas Process status which will help googlers real time to track the progress of each task and updates for enqueue of volume .

Impact : Dashboard was shared across 26 Project Areas of engineers to keep a track of progress of data labeling done by human raters by week on week updates and forecast for meeting the timelines as expected.

 **Data analysis and automation :**

* **QSR Indian Dining Food Chain Dataset Analysis:: Python Colab:** This Analysis was done to display the offers given by food chains online/offline. The aim of this analysis was to display the insights to the clients.
* **Enqueue & Dequeue on CNS tool/X20/GCS/ :: Linux Bash Shell Scripting or python scripts :** Created a Python Scripts to dequeue plugin data into colab and structuring data using sql and pandas then sharing insights with clients. Also enqueuing through linux bash scripts to enqueue in plugins as per the requirement and storing the output file into cns or x20 or gcs for sharing in csv formats.
* **Social Perception of People NLP :: Python Colab :** Label the regard (i.e Social perception of people) and sentiment (langage polarity) of sentence snippets. Each label has 5-6 categories ( Positive/negative/neutral/Both/Neither)
* Used Regular expression to clean the regard text data
* Applied Bag-of-Words is a method to represent text into numerical features. also using TF-IDF by penalizing the common words by assigning them lower weights.
* Mapping the Coherent and clarity columns by Categorical Numeric.
* Filling the missing values by the ffill method.
* Visualizing data using Python colab and presenting insights of analysis to the clients.
* **Text To Emotions :: Python Colab :** Given the Data in Text the task was to display the sentiments of the sentence using Emoji as output resulting images.

.

* Used Regular expression to clean the regard text data
* Removing stop words and targeting keywords which can display appropriate emoji for the text expression.
* Display of top 20 words influencing the certain emoji.
* Visualizing data using Python colab and presenting insights of analysis to the clients.
* **Custom Functions for plugins:: Google Cloud :** Working with protocol buffers to get structured data as requested by clients and also to optimize operation tasks such as performance optimization for human raters.

**Associate Analyst, Deloitte (C2H), Hyderabad (Feb-2018- Feb-2019)**

Part of Reporting & Analytics Team for WorkForce management as an analyst working extensively on SQL,Tableau,Excel.

* Automated reports using Tableau workbooks and also used Python libraries (Matplotlib) for Presentations.
* Gathered functional and non-functional client requirements in order to optimize the design of BI deliverables (reports, dashboards, alerts, visualizations) and Developed interactive graphics and data visualizations of large structured data in browser-friendly formats.
* Generating an automated Scatter Plot for Practitioner which compares with peers of that particular offerings using Tableau.
* Data Visualization and Automation of QBR reports for more than 15 Offerings using Tableau and Excel.
* Used Sql to generate Reports and integration with tableau. Also used Sql for automation of reports which requires manual updating daily.
* Created an automated userform which captures and consolidates cell values into cells for further analysis.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Projects and competitions**

Taken Part in multiple Machine learning Competitions and Data science Projects and Competitions:

**Twitter Sentiment Analysis :** The objective was to detect hate speech in tweets.. So, the task is to classify racist or sexist tweets from other tweets.

* Used Regular expression to clean the tweets data
* Applied Bag-of-Words is a method to represent text into numerical features. also using TF\-IDF by penalizing the common words by assigning them lower weights.
* Applied Random forest and predicted the value of Test data set with accuracy of 96 percent.

**Foreclosure Prediction (Ranked: Top 3 % in Competition) :** Given the Customer demographic, customer transaction, foreclosed customer the objective was to predict the probability of foreclosure for each of the data points.

* Merging all the data by customer demographics transaction and foreclosed customer and cleaning data by imputing missing values.
* Applying one hot encoding to relevant columns to get better insights from data.
* Applied Multinomial Naïve base and predicted the dataset.

**Incedo Data Science ( Ranked : Top 10%) :**.he objective was to Predict the addiction probability to drugs. Here the addiction prevention society are doing a survey and asking questions on such bachelors who are addicted to smoking or other drugs, and make strategies in order to prevent or control consumption of such things.

* Changing the categorical variables into binary numerical numbers using label encoder
* Splitting the Geographical location to Latitude and longitude with separate columns and filling the missing values mean of the column.
* Changing Data types of the new Feature engineering columns
* Extracting Date time and separating it in year, month, day : Hour, minutes ,seconds
* Applied Random forest regressor and predicted the value of Greater\_Risk\_Probability for Test data set.