RAKESH CHOUDHURY

Boston, MA | rakeshchoudhury6@gmail.com | 8572108264 | https://linkedin.com/in/rakesh--choudhury
github.com/rakesh-choudhury | https://rakesh-choudhury.github.io

EDUCATION

Northeastern University,
Master of Science in Information Systems *Expected: May 2021*

Kalinga Institute of Industrial Technology,
Bachelor of Technology, Electronics and Telecommunication *May 2011-May 2015*

WORK EXPERIENCE

Serole Technologies, India*- Data Analyst* *February 2018-July 2019*

- Built classification models leveraging random forest risk model with **0.81 F1 score** to segregate customers on their likelihood to payback for the Insurance product.
- Developed **ETL** data pipelines running **SQL stored procedure** pulling 100 GB data into target data marts for quantitative analysis following audit procedure.
- Led team in improving database and Query performance by nearly 50% using **SQL indexing technique**.
- Delivered **Tableau** report from 1000+ parameters of customer and finance data to executive team to improve customer acquisition and retention, increased sales pipeline.

Accenture, India*- Data Analyst* *June 2015-January 2018*

- Initiated SQL Query automation script using python for replication to target 500+ data mart.
- Led team to design **star schema** dimension model decreasing Query time by 50% adhering to business KPIs.
- Designed paralleled daily **ETL** night jobs with slowly changing dimension(**SCD**) type 2 technique pushing to **data warehouse** tables at **9x speed** for business analysis on **PowerBI**.
- Conducted **scrum** activities for team of 3 in **jira** environment, updating sprint backlog, and handling bugs and tasks.
- Won ace award for helping client save **$160,000** per year for its highest paying customer.

TECHNICAL SKILLS

Programming skills:	SQL, R, Python, Java, CSS, JavaScript
Libraries:	Numpy, Pandas, Tensorflow, Pytorch, Matplotlib, Plotly(DASH)
Version Control:	GitHub, Git, Bitbucket, LINUX(Bash)
Data Visualization tools:	Tableau, Power BI, Salesforce (Einstein Analytics), Flourish
Cloud:	AWS (EC2, S3, RDS, Redshift), Heroku
Tools:	Jira, Microsoft office suite, G-suite, Alteryx (ETL), Talend (ETL), Snowflake
Database:	SQL Server, Mongo DB, SAP HANA, PostgreSQL, MySQL, Oracle
Framework:	Django, Flask, Streamlit
Certifications:	Google Analytics for Beginners[Link], Advanced Google Analytics[Link]

ACADEMIC PROJECTS

IMDB Data Warehousing Project:(Amazon S3, PowerBI, Data Model, Alteryx, Dimensional model, BigQuery)

- Brewed Talend **ETL** jobs performing data cleaning with **regex** and data type conversions operations while loading 20 GB of flat file data into an **Oracle Data Warehouse**.
- Optimised all jobs ingesting data into Oracle database tables from data stored in **AWS redshift** saving up to **3 GB** of memory.
- Loaded data into **PowerBI** with interactive dashboard for 100+ attributes, measures and calculations resulting visualization of data and findings.

Airport Database Management System:(*Tableau, RDBMS, Business Intelligence, AWS RDS, MSSQL,Talend*)

- Plumbed Talend data pipeline to load dimensional model data into **SQL server** from **Amazon RDS** with reduction in memory consumption by 1.5 GB.
- Manipulated 4 GB of data using **SQL** joins, functions, triggers, procedures, error handling and pivot table.
- Integrated data using **SQL views** as different data marts to be consumed into **Tableau** dashboard with 15+ parameters to generate insights on Airport operations

Recommendation system A/B testing:(Machine Learning, Python, Heroku, Streamlit, Numpy, Pandas, Seaborn, Matplotlib)

- Utilized Tensorflow RBM (Restricted Boltzmann Machine) **machine learning** model in python for training **neural networks** feeding product ratings, historical behavior, Retail Channel as key parameters achieving an accuracy of 79.2% in predicting 1-5 star ratings.
- Productionised **machine learning** application leveraging **streamlit** on **Heroku** for a recommendation system; placing products chronologically that engages users and helps in shooting sales upto **250%**.
- Operated on a large scale **recommendation system** to improve market strategy, customer experience, and revenue bump by 30%.

Brand Promotion on YouTube Channels:(NLP, Logistic regression, k-means, Amazon EC2, Flask, Statistical Methods)

- Handled 4 GB dataset with **python** classifying YouTube Channels with popularity score on likes, dislikes, views via **scikit learn's k-means** and **logistic regression** model.
- Computed trend metric with popularity classification points and **NLP**'s sentiment scores resulting in **225%** return on investment as part of brand influencing.
- Integrated python code on **Flask** web application and deployed on **AWS EC2** instance.