# RAKESH CHOUDHURY

Boston, MA | rakeshchoudhury6@gmail.com | 857.210.8264 | https://linkedin.com/in/rakesh--choudhury
github.com/rakesh-choudhury | https://rakesh-choudhury.github.io

## EDUCATION

***Northeastern University,***
*Master of Science in Information Systems* *September 2019-May2021*
***Kalinga Institute of Industrial Technology,***
*Bachelor of Technology, Electronics and Telecommunication* *May 2011-May 2015*

## TECHNICAL SKILLS

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

## WORK EXPERIENCE

***Serole Technologies, India- Data Analyst*** *April 2018-December 2018*

- Built classification models leveraging random forest 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.
- Helped digital marketing team save on ad expenditure by approximately **$5000** per each policy ad through **A/B testing** analysis on users data.
- Delivered **PowerBI** visualisation helping company make decisions to improve customer acquisition and retention, enhancing sales pipeline resulting in **100%** revenue increase.

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

- Initiated SQL Query automation script using python for replication to 10+ target data mart.
- Designed **ETL** jobs with slowly changing dimension(**SCD**) type 2 technique while loading **data lake** tables at **9x speed** for business reports on **Tableau**.
- Optimized database and Query performance by nearly 3x faster using **SQL indexing technique**.
- Won ace award for helping client save **$160,000** per year for its highest paying customer.

## ACADEMIC PROJECTS

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

- Created Alteryx **ETL** performing data cleansing operations while loading 20 GB of data from **AWS S3** into **Oracle Data Warehouse**.
- Wrote ETL workflows integrating **facts with dimensions** into Oracle database saving upto 3GB of memory.
- Visualised data in **PowerBI** with interactive dashboard for 100+ attributes, measures and calculations of data and findings.

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

- Constructed **Talend** data pipeline to load dimensional model data into **MicrosoftSQL server** from **Amazon RDS** reducing memory consumption by 1.5 GB.
- Manipulated 4 GB of data using **SQL** query joins, functions, triggers, procedures, error handling, CTE and pivot table to create reports for business consumption.
- Integrated data from different source tables into **SQL views** to be consumed in **Tableau** dashboard with 15+ parameters to generate insights on Airport operations.

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

- Utilized Tensorflow RBM (Restricted Boltzmann Machine) **machine learning** model for training **neural networks** feeding product ratings, historical behavior, Retail Channel as key parameters achieving an accuracy of 79.2%.
- Deployed **streamlit** application on **Heroku** for a recommendation system; placing products chronologically that engages users and helps in shooting sales upto **250%**.

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

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