RAKESH CHOUDHURY

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

SUMMARY

Experienced Data Engineer delivering solutions in scrum-based agile environments, helping clients develop analytical reports with dashboards for business analysis. I am a data pirate who loves hunting for the most valuable data to bring out some great insights with visualizations.

EDUCATION

Northeastern University,
Master of Science in Information Systems — *Expected: May 2021*
Courses: Data Science Management Tools, Application Engineering Development, Program Structure Algorithm, Algorithmic Digital Marketing, Data Management Database Design, Designing Data Architecture Business Intelligence

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

WORK EXPERIENCE

Serole Technologies, India*- *Data Engineer — *February 2018-July 2019*

- Massaged data using SQL reporting with classification of more than 1 million user's/customer's eligibility for insurance product.
- Mined approximately 100 gigabytes of risk management data for quantitative analysis following audit procedure to determine company's loss or gain based on product damage, lifespan, and other parameters of products.
- Trained new team about project's technical and business aspects to help scale deployments and process reports by 10%.
- Developed stored procedures to build and optimise data pipelines for ETL processes freeing up more or less 30 gigabytes of memory space to ingest transactional data into data lake tables to be consumed by data science and reporting teams.
- Led team in solving and improving complicated database and query performance by nearly 50% adhering to business best practices while working with a small team with varying skill sets.

Accenture, India*- *Data Engineer — *June 2015-January 2018*

- Conducted scrum activities for team of 3 in assigning and updating tasks and stories in jira environment, updating sprint backlog, and handling bugs and tasks.
- Initiated automation project collaborating with team members where SQL script generated using python for 70+ systems for a change request. This helped client save approximately $160,000 per year for 20 change requests.
- Analysed client's business requirement and SQL views report using statistical methodologies to track post production support ticketing system performance and helped improve effectiveness of communication to customers by 30%.
- Envisioned the data on PowerBI dashboard stories with KPIs that depicts department that need improvement.
- Won ace award for database performance optimisation and automation for its highest paying customer.

TECHNICAL SKILLS

Programming skills:	SQL, R, Python (Numpy, Pandas, Tensorflow, Pytorch), Java, Bash, CSS, JavaScript
Operating Systems:	Mac OS, Linux, Windows
Version Control:	GitHub, Git, Bitbucket
Data Visualization tools:	Plotly(DASH), Matplotlib, Salesforce (Einstein Analytics), Flourish, Tableau, Power BI
Cloud:	Heroku, AWS (EC2, S3), HPC(High Performance Computing) cluster
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

Recommendation system A/B testing: — *September 2020-January 2021*

- Utilized Tensorflow RBM (Restricted Boltzmann Machine) machine learning model in python for training neural networks feeding product ratings 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 product usage, customer experience, and revenue bump by 30%.

Brand Promotion on YouTube Channels — *May 2020-August 2020*

- Classified YouTube Channels based on data analysis of popularity, likes, dislikes, views via scikit learn's k-means and logistic regression model to analyse dataset with python.
- Computed trend metric extracting parameters from arima time series forecasting and NLP's sentiment scores to monitor the brand sentiment and potential market reactions.
- Incorporated above parameters to formulate return on investment resulting up to 225% on certain channels.
- Integrated python code on Flask web application and deployed on AWS EC2 cloud server.

IMDB Data Warehousing Project: — *September 2019-December 2019*

- Brewed Talend ETL jobs performing different regex and data type conversions operations while loading flat file data into an Oracle Data Warehouse.
- Parallelized all jobs in master job to optimize performance while feeding 15 GB data into PostgreSQL database tables from data stored in Amazon S3.
- Reconfigured tables into dimensional model designed in ER Studio, enhancing the query runtime by roughly -50% and enabling faster reporting and analyzing capabilities.
- Loaded data into PowerBI with interactive dashboard for different attributes, measures and calculations resulting visualization of data and findings.