RAKESH CHOUDHURY

LinkedIn | rakeshchoudhury6@gmail.com | Website | GitHub | +1 857-210-8264 | Boston, MA

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 user's/customer's eligibility for insurance
- Mined risk management data for quantitative analysis 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

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
- Initiated automation project collaborating with team members where **SQL script** was generated using **python** for 70+ systems for a change request. This helped client save approximately **$160,000 per year** for 20 change requests
- Developed stored procedures to build data pipelines for ELT processes to **ingest** transactional data into **data lake** tables
- Analyzed client's business requirement and made SQL views report using statistical methodologies to track post production support ticketing system performance
- Envisioned the data on **PowerBI** dashboard stories with KPIs that depicts department that need improvement
- Led team in solving complicated **database performance** issue adhering to **business best practices**
- Won **ace award** for innovative database performance optimization

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 suit, G-suit, Alteryx (ETL), Talend (ETL), Snowflake, Microsoft Sharepoint
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 project: [Project Link]   *September 2020-January 2021*
- Utilized **Tensorflow** RBM (Restricted Boltzmann Machine) machine learning model in python for training neural networks feeding product ratings as key parameter achieving an accuracy of 79.2%
- 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%

Brand Promotion on YouTube Channels: [Project Link]   *May 2020-August 2020*
- Classified YouTube Channels based on popularity, likes, dislikes, views via scikit learn's k-means and logistic regression model
- Computed trend metric extracting parameters from **arima** time series forecasting and NLP's sentiment scores
- Incorporated above parameters to formulate **return on investment resulting up to 225%** on certain channels
- Integrated python code on **Flask** web application and deployed it on **AWS EC2** cloud

Airport Database Management System: [Project Link]   *January 2020-April 2020*
- Fed data to **SQL server database** from flat file anchoring **ETL pipelining** tool **Alteryx** by following **star schema data models**
- Assembled **functions**, **triggers**, **procedures** and **error handling** to maintain data integrity and consistency in relational **SQL server** database
- Leveraged **pivot table** functionality to create clear readable reports on tabular output
- Worked on **statistical analysis** of customer centric data on **Tableau** dashboards and plotted flight routes on map to check the busyness of airport

IMDB Data Warehousing Project: *September 2019-December 2019*
- Brewed **Talend** ETL jobs performing different regex and data type conversions operations while loading flat file
- Configured connections in **Talend** to load data to staging tables in target **PostgreSQL** database
- Parallelized all jobs in master job to **optimize** performance while feeding 15 GB data into **PostgreSQL** database tables
- Reconfigured tables into **dimensional** model designed in **ER Studio**, enhancing the query runtime by roughly -50%
- Loaded data into **PowerBI** with interactive dashboard for different attributes, measures and calculations resulting visualization of data and findings