Ganesh Raj K

◆ ganesh_012@outlook.com ◆ 848 313 8525 ◆ linkedin.com/in/ganeshrajk ◆ github.com/ganeshraj-k

EDUCATION:

- **MS in Data Science**, Rutgers University | CGPA – 3.75 May 2024
- **BTech in Computer Science and Engineering**, Indian Institute of Technology Indore May 2019

EXPERIENCE:

Rutgers UCM, *Data Analyst* **Feb 2023 – Present, New Brunswick**

- Harnessed learning data from canvas Api to predict new student course outcomes/CGPA grade using a classification model which scored a 78 percent success rate in classifying new students.
- The dataset consisted of 12 diverse features including demographic academic, behavioral, parent participation for the predictive modeling.
- Extracted data using canvas Api, preprocessed using python, and used ANN for the classification model.

Deloitte Consulting, Data Analyst **June 2019 – Jan 2022, Bengaluru**

Medical Data NER:

- Enhanced query speed for a medical record database with over 2 million records for a major healthcare client.
- Used Amazon Comprehend with Python to perform Named Entity Recognition on the DynamoDB dataset.
- Added recognized entities as tags using AWS Glue for the ETL process.

Banking:

- Mitigated the lockdown-induced customer churn by constructing a multivariate logistic regression model in an Agile environment to Identified churn-prone customers and key contributing factors.
- Performed EDA using matplotlib and communicated results to stakeholders. helped them target marketing accordingly resulting in a 30% churn reduction in the next quarter.

***Key Achievement**:* Was recognized with Applause award twice for my client centric work approach and timely deliverables.

MAQ Software, Data Engineer **May 2018 – July 2018, Hyderabad**

- Established an ETL pipeline using SQL Server Management Studio and SSIS, consolidating three large Azure data marts with over 2 million records into one. Developed triggers and stored procedures in place to identify inconsistencies during the transfer and maintain data integrity.

PROJECTS:

Twitter Search: [github]

- Designed a web application with a local cache of 200 trending tweets, leveraging a combination of Postgres (relational) and MongoDB (non-relational) to query a dataset of about 120,000 tweets from 13,000 users.
- Applied NLP techniques for efficient search, including synonym search and Levenshtein distance, and managed API requests with Flask

2024 Travelers Insurance Analytics University Contest: [github]

- Conducted Tweedie regression on a zero-inflated dataset of over 29,000 records, fine-tuned parameters using grid search, and assessed model efficacy with the Gini index. This systematic approach secured a third-place finish among 200+ teams.

SKILLS

- Programming Languages: Python, R
- Machine Learning Libraries and Frame works:: PyTorch, scikit-learn, pandas, numpy
- Cloud: AWS, DynamoDB, Glue, EC2, Sage Maker, IAM, S3
- Office: Excel, PowerPoint, VBA
- Data Visualization: Tableau, Power BI, Matplotlib, Seaborn, Plotly

CERTIFICATIONS

- AWS Machine Learning Specialist
- AWS Cloud Practitioner