|  |  |  |
| --- | --- | --- |
| New Brunswick, NJ  (848) 313 8525  ganesh\_012@outlook.com | **GANESH RAJ K** | Linkedin.com/ganeshrajk  Github.com/ganeshraj-k |

**E****DUCATION**

* ***Master’s in Data Science,*** *Rutgers University New Brunswick May 2024*
* ***Bachelor’s in Computer Science,*** *Indian Institute of Technology Indore May 2019*

**TECHNICAL SKILLS**

* **Programming Languages:** Python, R
* **Libraries and Frameworks:** Git,  Spark, PyTorch, TensorFlow, MongoDB, Postgres SQL
* **AWS:** lambda, S3, CloudWatch, Sage maker, IAM, EC2
* **AWS certified Cloud practitioner**
* **AWS certified Machine Learning specialist**

**EXPERIENCE**

|  |  |
| --- | --- |
| **Unit Computing Specialist (student worker), Rutgers UCM** | **June 2019 – Jan 2022, Bangalore** |

* Developed Python scripts for Excel data cleaning, reducing manual work by 30+ hours quarterly.
* Created Tableau dashboards for budget analysis and IT ticket trends, identifying common issues to support decision-making.

|  |  |
| --- | --- |
| **Business Analyst, Deloitte** | **June 2019 – Jan 2022, Bangalore** |

* Reduced banking customer churn by 33% during lockdown by identifying churn customers and key factors using logistic regression, while helping the client target strategies toward online features.
* Reduced manual effort by over 35 hours each week in identifying boats on a dock. Developed a change detection system using OpenCV and Mask R-CNN, and automated the process using AWS Lambda and CloudWatch
* Assisted a large restaurant chain in achieving an over 80% increase in take-away orders during COVID-19. We segmented customers based on demographics and risk tolerance using DBSCAN clustering to tailor their marketing approach. Additionally, I created a Tableau dashboard to present understandable insights.
* Saved over $100k in internal HR work hours by developing an employee search system using R Shiny for my vertical with over 800 employees. Leveraged NLP in R to extract skills from resumes and managed a real-time MongoDB database for up-to-date tracking
* Improved medical record querying speed by over 110% for more than 2 million records. Achieved this by establishing an ETL pipeline using AWS Glue and leveraging Amazon Comprehend for Named Entity Recognition. The identified entities were used as tags for better indexing.

|  |  |
| --- | --- |
| **Database Management Intern, MAQ Software** | **June 2019 – Jan 2022, Bangalore** |

* Implemented an ETL pipeline in SQL Server Management Studio and SSIS to clean, structure, and consolidate three Data Marts into one. Also developed triggers and stored procedures to identify and resolve errors and inconsistencies in tables during the transfer process.

**PROJECTS**

***Chatbot model with a personality:***

* Developed a generative AI (Gen AI) seq2seq model to replicate Chandler Bing’s dialogue style from “Friends,” utilizing an extensive dataset of 8,700 dialogues. The model, featuring a 2-layer LSTM network with a dropout layer, achieved a BLEU score of 0.63.
* Preprocessed the data using Python and created Tableau dashboards to analyze the show’s data, such as average season ratings, episode rankings, character interactions, and dialogue frequency, which informed further model development and data understanding.

***2024 Travelers Insurance Analytics University Contest:***

* Performed Tweedie regression on a zero-inflated dataset with over 29,000 records, fine-tuning parameters via grid search and assessing model efficacy with the Gini index.
* Executed comprehensive EDA and data visualizations to discern distributions of individual attributes, which informed model selection. This methodical approach garnered us third place among 200+ teams and recognition on campus for precise insurance claim predictions. precise insurance claim predictions.

***Twitter Search****:*

* Created a web application for querying a dataset of ~120,000 tweets from 13,000 users, utilizing a combination of Postgres (relational) and MongoDB (non-relational), with a local cache of 200 trending tweets.
* Applied NLP techniques for efficient search, including synonym search and Levenshtein distance, and managed API requests and the web app using Flask.

***Ear image recognition***

* Implemented image recognition on ear images using a CNN model with SIFT features for biometric authentication. Captured and preprocessed high-resolution images of students, achieving an accuracy of 91% and a precision of 73%.

**VOLUNTEERING**

* **Club Head, AVANA**: Led AVANA, the campus social welfare club at IIT Indore, spearheading initiatives such as cleanliness drives, weekend teaching workshops for underprivileged children, and blood donation drives.
* **Volunteer Team Lead, Breathe India:** During the COVID-19 second wave, I facilitated access to medical resources by daily updating our website with information on available beds, nebulizers, and medicines, and promptly connected individuals to the necessary healthcare services.