

EDUCATION

Texas A&M University <i>Master of Science in Data Science</i>	College Station, TX <i>Aug. 2024 – Present</i>
SRM Institute of Science and Technology <i>Bachelor of Technology in Computer Science Engineering</i>	Chennai, India <i>Jul. 2018 – May 2022</i>

EXPERIENCE

Student Research Assistant <i>School of Public Health, Texas A&M University</i>	Oct. 2024 – Present <i>College Station, TX</i>
<ul style="list-style-type: none">Conducted statistical analysis on Medicaid data to derive actionable insights using Python and SAS.Automated graph generation for project presentations using R, reducing processing time by 30%.Performed hypothesis testing using T-tests and Pearson correlation to identify relationships between target columns, and utilized Q-Q plots to analyze distribution similarities across datasets over multiple years.	
Data Analyst <i>Draup Business Solutions</i>	Dec. 2022 – Jun. 2024 <i>Bangalore, India</i>
<ul style="list-style-type: none">Implemented scalable and robust ETL data injection pipelines with PySpark and SQL in AWS EMR jupyter notebooks, enhancing reliability and resulted in a 35% improvement in end-to-end data integrity.Led a cross-functional initiative to architect a dashboard, processing 200 million data checks through Airflow to generate 100+ exceptions; rapid resolution of production issues.Integrated AWS Lambda with S3 and DynamoDB to support ad-hoc client data request.	
Data Analyst <i>HighRadius Corporation</i>	Jul. 2022 – Nov. 2022 <i>Hyderabad, India</i>
<ul style="list-style-type: none">Transformed data gathering phase by simplifying data extraction and preprocessing steps with Python and SQL, decreasing total time for analysis by 50%.Designed a Power BI dashboard to share insights with stakeholders, reducing decision-making time by 40%.	
Data Scientist <i>HighRadius Corporation</i>	Aug. 2021 – Jun. 2022 <i>Hyderabad, India</i>
<ul style="list-style-type: none">Collaborated with multiple Fortune 500 CPG companies to facilitate AR work distributions utilizing time series data in Python, reducing manual efforts by 4 times.Created predictive models for customer payment date patterns leveraging machine learning regression models, Bagging, and boosting algorithms like LightGBM, resulting in a 70% increase in model accuracy.Optimized deployed ML models, boosting automation efficiency by 25% and achieving 35% revenue savings	

PROJECTS

Metastatic Cancer Detection <i>Python, CNN, Pytorch, Git/Github, Neural Networks</i>	Nov. 2024 – Dec. 2024
<ul style="list-style-type: none">Integrated deep learning model using CNNs to classify metastases in histopathological images from the PatchCamelyon (PCam) dataset.Applied data augmentation and batch processing to enhance model generalization and training efficiency.Achieved an F1 score of 0.8768, demonstrating high accuracy in automated cancer metastasis detection.	
Metro Interstate Traffic Volume <i>Python, Statistical Techniques, Model Development</i>	Nov. 2024 – Dec. 2024
<ul style="list-style-type: none">Built a traffic congestion model using scikit-learn, with feature scaling, one-hot encoding, and time-series analysis.Applied Random Forest, Lasso, Ridge, Linear, and Polynomial Regression with cross-validation.Used pandas, NumPy, Matplotlib, and Seaborn for data analysis; tuned hyperparameters with GridSearchCV.Evaluated with RMSE and R-squared; Polynomial Regression showed the best accuracy.	

SKILLS

Languages & Framework: Python, R, SAS, Pandas, NumPy, Matplotlib, Scikit-learn
Database: Postgres, MySQL, MongoDB, DynamoDB, Snowflake
Cloud Computing: AWS Suite (EMR, S3, EC2, Glue, Lambda), Spark, PySpark
Tools/Platforms & Interest: Git/GitHub, CI/CD, LLMs, Power BI, Tableau, Data Analytics, Pattern Analysis