

Aditi Reddy Doma

ad7977@rit.edu | 917-337-7045 | [linkedin.com/in/aditireddydoma](https://www.linkedin.com/in/aditireddydoma) | [GitHub](https://github.com)

EDUCATION

Rochester Institute of Technology

MS in Data Science (GPA: 4.0)

Rochester, NY

Expected: December 2026

- Relevant Coursework: Software Construction, Applied Statistics, Foundations of Data Science, Database Design Implementation, Data Visualization

Mahindra University

Bachelor of Technology in Artificial Intelligence (GPA: 3.2)

Hyderabad, India

Graduation Date: May 2024

- Relevant Coursework: Data Structures and Algorithms, Machine Learning, Big data, Deep Learning, Artificial Neural Networks, Data Mining and Warehousing, Risk Management, High Performance Computing

WORK EXPERIENCE

Neovatic Technologies

Business Analysis Intern

Hyderabad, India

Feb 2024 – June 2024

- Documented client requirements by collaborating with **15+ cross-functional** team members and **attending client meetings**.
- Extracted **actionable insights** from **20,000+ rows** of patient data using **MS Excel**, then designed a low-code application in **Mendix**, leading to improved patient onboarding and streamlined appointment scheduling.
- Presented the Mendix app functionality to technical and non-technical stakeholders using clear **MS PowerPoint** decks.

Cloud4C

Machine Learning & Data Analysis Intern

Hyderabad, India

June 2023 - August 2023

- Achieved **Natural Language Processing (NLP) accuracy above 90%** by developing a sentiment analysis product using **BERT** through transfer learning and fine-tuning on customer raised tickets.
- Optimized training pipelines in **TensorFlow** using **dynamic learning rates**, **mixed-precision training** and **gradient clipping** to improve training efficiency and model stability.
- Identified the **top 5** recurring customer concerns by applying **K-Means clustering** on BERT embeddings.

PROJECTS

Portfolio Allocation Model ([GitHub](#))

- Researched and scraped **5 years of stock data** for 9 companies across 3 sectors using Yahoo Finance, and optimized portfolio allocation through **mean-variance modeling** using **Pyomo, NumPy, and Pandas**.
- Identified Buy & Hold as the optimal investment strategy with **15% ROI** by performing **statistical analysis** of risk-return trade-off assessments using **Monte Carlo simulation** in **Python**.

Strategic Insights for Mitron Bank's New Credit Card Line ([GitHub](#))

- Developed **3 Power BI dashboards** on income analysis, expenditure trend and customer segmentation by demographics.
- Designed a web like interface with button based navigation and **statistical charts**.

Energy Consumption Predictor ([GitHub](#))

- Forecasted hourly energy consumption with **92.54% accuracy** using **time-series Linear Regression** in R for 5000+ homes.
- Built a **Shiny dashboard** to show how temperature, time and house attributes **influence future energy consumption**.
- Visualized county level energy hotspots using **ggplot2** and **plotly**, highlighting high consumption zones for energy insights.

Cloud Infrastructure ETL Pipeline for YouTube Analytics Data Engineering ([GitHub](#))

- Built a cloud-native ETL pipeline using **AWS** and **PySpark** to analyze YouTube trends across countries, cleaning and converting **170K+** semi-structured records to partitioned Parquet format in Amazon S3.
- Developed **3+ AWS Glue jobs** and **Lambda** triggers for regional ingestion, schema mapping and country level partitioning.
- Queried datasets in **Athena** and visualized regional category trends to enable content performance insights using SQL.

Covid Data Exploration and Visualization ([GitHub](#))

- Cleaned data using **Excel**, automated updates via **Power Automate** and conducted **exploratory analysis** with **SQL queries**.
- Showcased COVID-19 trends and vaccination impact in **Tableau** using **multiple visualizations** including charts, maps.

Diabetes Disease Prediction ([GitHub](#))

- Trained 7 machine learning **classifiers** in **Python** (**Logistic Regression, SVM, Gradient Boosting, XGBoost, Decision Tree, KNN, Random Forest**) to predict diabetes based on patient health metrics.
- Applied **one-hot encoding** and **MinMax normalization** for data cleaning and split data into **75/25 train-test**.
- Found Gradient Boosting most effective with **91.45%** prediction accuracy and **root mean squared error (RMSE)** of **0.39**.

SKILLS

- **Programming & Libraries:** Python, Java, R, MySQL, Pandas, NumPy, Scikit-Learn, Keras, TensorFlow, PyTorch, Git
- **Data Analysis Tools:** Tableau, PowerBI, Seaborn, Matlab, Excel, Word, Outlook, PowerPoint, Teams, SharePoint, AWS
- **Machine Learning:** Supervised & Unsupervised Learning, NLP, A/B Testing, Mathematical concepts, Statistical Analysis