RUTHVIKA REDDY TANGIRALA

Boston, MA (Open to Relocation) | (617) 602-0966 | ruthvikareddytangirala20@gmail.com

LinkedIn | GitHub| Portfolio

EDUCATION

Northeastern University, Boston, MA **July 2025**

Master of Science in Data Science and Analytics GPA: 3.92

Coursework: Database Management, Data Mining, Machine Learning, Machine Learning Operations, Data Warehousing

TECHNICAL SKILLS

Programming languages: Python, R, Structured Query Language (SQL), Git

Data Science: Machine Learning, GenAI, Feature Engineering, Statistical Modeling, PyTorch, TensorFlow, Scikit-learn

Database & BI: MySQL, PostgreSQL, MongoDB, Oracle, Pinecone, Snowflake, Redshift, Data Modeling, Excel

Data Engineering: Apache Airflow, ETL/ELT, PySpark, Docker, Kubernetes, MLflow, CI/CD, DVC

Cloud: AWS (S3, SageMaker, EC2, Lambda, Athena, Glue, Bedrock), GCP, Azure (Data Lake, Databricks), Salesforce

Certification: IBM Data Science Professional Certificate, AWS Certified Cloud Practitioner

WORK EXPERIENCE

Mercedes-Benz RDNA, Farmington, MI Sep 2024 – May 2025

Data Scientist | Python, SQL, Power BI, Azure, GenAI, Automobiles

- Engineered a GenAI-powered pipeline using proprietary LLMs to extract, translate (German to English), and summarize 20+ years of meeting notes, enabling semantic search and reducing manual lookup time by 90%
- Developed a Retrieval-Augmented Generation (RAG) system powered by AI agents and LLMs to perform reasoning over internal data sources, cutting manual analysis effort by over 50%, and enabling contextual, multi-source answers to complex user queries
- Designed a Power BI dashboard with a Microsoft SQL Server backend to track CARB approval stages and KPIs for OBD groups, with a custom Python GUI for real-time data updates, streamlining compliance visibility across engineering teams
- Automated a configurable data ingestion and transformation pipeline using a custom Python GUI and backend scripts, automating the flow of multi-source vehicle test data (INCA, GST, Emissions) into Microsoft SQL Server for business intelligence reporting
- Collaborated with cross-functional and globally distributed teams to improve documentation standards and streamline Agile workflows, ensuring the reliability of automated processes

IBM India Private Limited, Hyderabad, India Feb 2022 – August 2023

Data Scientist | SQL, Salesforce CRM, Healthcare

- Executed complex SQL queries to extract critical object data and data records, facilitating enhanced data-driven decision-making and significantly improving system performance
- Constructed interactive dashboard through data ingestion using ServiceNow data (incidents and RITM records), to monitor production-level bugs and issues, providing weekly and monthly insights for proactive resolution and optimization
- Spearheaded development of an efficient Chatbot for 'Blue Cross Blue Shield' in Salesforce Cloud, using over 50 tailored instances per question for training and integrating machine learning for quick, accurate customer responses and streamlined case creation

PROJECTS

INCOME CLASSIFIER BASED ON THE DEMOGRAPHIC DATA| Python, Analysis, Machine Learning

- Implemented a spectrum of machine learning algorithms, including KNN, Decision Trees, Random Forest, Logistic Regression, SVM, and Neural Networks, attaining an accuracy of 86% and a recall of 72%
- Enhanced model accuracy by up to 3% through advanced data preprocessing, including imputation of missing values, deduplication, data type normalization, and multicollinearity reduction using Variance Inflation Factor (VIF)

Hospital Database Management System| SQL, Analysis, AWS, Healthcare

- Designed and deployed a normalized healthcare database schema on Amazon RDS (SQL Server) to manage patient, staff, pharmacy, insurance, and appointment data across 17 interrelated tables with defined 1:1 and 1:N relationships
- Implemented stored procedures, triggers, and constraints in RDS SQL Server to enforce referential integrity, automate calculations, and ensure high data quality for analytics and operational reporting

Personal Trainer AI Agent | Python, GCP, RAG, LangChain, LangGraph, PostgreSQL, Cloud Deployment

- Built a GenAI-powered fitness assistant using RAG over curated content (PDFs, websites) stored in Pinecone vector database, enabling personalized, goal-based workout recommendations with real-time context awareness based on multi-agent system
- Orchestrated end-to-end pipelines for data ingestion, semantic search, and deployment pipelines using Docker, Airflow, MLflow, and GitHub Actions on GCP, achieving rapid iteration, robust monitoring, and zero manual deployment