Sai Sriram Uppada

Fairfax, VA 22030 | saisriramuppada596@gmail.com | +1 (704) 956-9712 | linkedin.com/in/saisriramuppada

Portfolio: saisriramuppada.com

Professional Summary

Data Analytics Engineering graduate holding AWS Certified Data Engineer (Associate) credential with hands-on experience architecting end-to-end data pipelines, machine learning models, and LLM-driven analytics systems. Proficient in Python, SQL, PySpark, and AWS cloud platforms, demonstrating practical expertise in ETL workflow design using SnapLogic, data preprocessing, feature engineering, and model evaluation across structured and unstructured data. Successfully delivered cloud-native solutions addressing real-world challenges, including operational data store design for eRebate's transactional data, intelligent resume–job matching systems leveraging multi-LLM inference, and crash-severity prediction models. Skilled in AWS services (S3, Glue, EMR, Aurora, Redshift), AI-ML frameworks, modern big data processing frameworks, and ETL technologies. Comfortable collaborating with databases, visualization tools, BI platforms, and Big Data technologies within Agile, research-driven environments. Actively seeking data engineering and analytics roles to architect scalable systems, develop robust data platforms, and drive measurable business impact through intelligent automation and AI-driven solutions.

Education

George Mason University, Fairfax, VA, Jan 2024 – Dec 2025
Master's in Data Analytics Engineering (GPA: 3.85/4.0)

Gayatri Vidya Parishad College of Engineering, Visakhapatnam, India August 2019 – May 2023
Bachelors in Information Technology

Experience

Graduate Research Assistant, George Mason University – Hybrid, Fairfax County Jan 2025 – July 2025

- Contributed as a junior research assistant on a project to design and evaluate an Operational Data Store (ODS) for eRebate's transactional data, focusing on how data pipelines and cloud databases can support reliable reporting and analytics.
- Assisted in building and testing ETL workflows using Python and SnapLogic, working with flat-file data (CSV/TSV) and helping transform it into structured formats suitable for storage in AWS Aurora PostgreSQL. Supported the team in organizing schemas and tables in Aurora PostgreSQL, and observed how design choices affect performance, scalability, and downstream analytical use cases.
- Helped prepare and review data flows between AWS S3, Aurora, Lambda, and Glue, gaining hands-on exposure to cloud-native data engineering patterns in a supervised research environment. Documented experiments, pipeline behavior, and data quality findings, contributing to internal notes and knowledge sharing for future enhancements of the ODS architecture.

All India Council for Technical Education Internship, NEAT – Remote, India March 2022 – May 2022
Data Analytics Intern

- Designed and implemented a voter fraud detection system using facial recognition with OpenCV, leveraging the Eigenface algorithm for image analysis and classification. Utilized Python, NumPy, Pandas, and Scikit-learn for data preprocessing, feature extraction, and model evaluation. Integrated real-time detection with a face recognition library for database matching and anomaly detection. Integrated with Jupyter Notebook for development and debugging.
- Hands-on experience with data analytics and machine learning frameworks, utilizing Python, OpenCV, face_recognition, NumPy, Pandas, Matplotlib, and Scikit-learn. Focused on data preprocessing, exploratory data analysis (EDA), machine learning (classification), image processing, and real-time data integration.
- During the internship, I gained proficiency in statistical analysis, supervised learning, real-time analytics, predictive modeling, and real-time data processing to address complex problems in digital security landscapes.

All India Council for Technical Education Internship, NEAT – Remote, India Oct 2021 – Dec 2021
AI - ML Intern

- Developed a stock price prediction system using Long Short-Term Memory (LSTM) for time series forecasting, leveraging Python, TensorFlow, Keras, and Streamlit. Implemented data preprocessing with Pandas and NumPy, model evaluation with Scikit-learn, and visualized results using Matplotlib. Focused on training and testing regression models to enhance prediction accuracy.
- The project involved training and testing machine learning models to identify patterns and predict stock price movements, focusing on regression models for accurate prediction.
- During the internship, I gained practical experience in Python programming, deep learning (LSTM), data preprocessing, and visualization using tools like NumPy, Pandas, Keras, and Matplotlib, while understanding key concepts of machine learning, regression analysis, and time series forecasting.

Projects

LLM-Based Resume–Job Matching and SOC Classification System Aug 2025 - Dec 2025

- Built an end-to-end LLM-based analytics pipeline in Python to classify resumes and job listings using SOC-2018 codes and generate Top-10 job recommendations per resume. Processed resume and job-listing text from CSV, PDF, and JSON sources, performing data cleaning, normalization, and PII removal to ensure consistent inputs for LLM inference.
- Integrated multiple LLM APIs (OpenAI GPT-4o, GPT-4o-mini, Anthropic Claude, Cohere, and Mistral) using standardized prompts to enable fair and repeatable model comparisons across tasks.
- Designed and tested three prompt tiers (baseline, intermediate, comprehensive) and evaluated model performance by comparing LLM-generated SOC codes against manually labeled ground truth, targeting 90% accuracy. Stored predictions and evaluation results in MongoDB, created accuracy visualizations, and collaborated on GitHub documentation and experiment reproducibility.

Crash Analytics: Predictive Modeling of Road Accident Severity Using Machine Learning Jan 2025 - May 2025

- Led a research-driven study leveraging a dataset of 132,000 global road accidents to investigate the interplay of driver behavior, environmental context, and temporal factors in determining accident severity. Applied advanced supervised learning algorithms including Random Forest, Logistic Regression, and XGBoost, achieving 80% classification accuracy and an F1-score of 0.80. Addressed class imbalance using SMOTE to improve minority class detection.
- Conducted rigorous exploratory data analysis and feature engineering using Pandas, NumPy, and Matplotlib, identifying key predictors such as driver alcohol level, poor visibility, and high-speed zones. Employed interpretability techniques, including SHAP, LIME, and Partial Dependency Plots (PDPs) to explain model behavior and extract actionable insights for real-world safety interventions and policy recommendations.
- Developed data pipelines and model prototypes in Python (scikit-learn, XGBoost) and deployed interactive visualizations using Streamlit for collaborative review and analysis. Strengthened research competencies in machine learning, human-environment systems analysis, and statistical reasoning to contribute to multidisciplinary research in transportation safety and public health.

Real Estate Price Prediction Using Socioeconomic Indicators Aug 2024 - Dec 2024

- Conducted a data-driven research study analyzing over two decades of property sale records across Connecticut to examine the influence of crime rates, school accessibility, healthcare availability, and employment statistics on real estate pricing. Employed Random Forest and XGBoost regression models to identify nonlinear interactions, achieving an R^2 of 0.90 and high predictive fidelity.
- Executed rigorous data preprocessing, including median imputation, ZIP code normalization, and time-series transformations, to enable accurate modeling of sale price dynamics. Integrated heterogeneous datasets using town-based joins and applied log transformation for skewness correction and model stabilization. Achieved an 80% success rate in design acceptance among target user groups.
- Performed feature importance and residual error analyses to extract interpretable patterns, revealing property crime as a dominant factor in housing devaluation. Visualized regional disparities and temporal trends using Tableau and Matplotlib to support urban policy and socioeconomic planning decisions.

Fashion Designing Using GANs Jan 2023 - Apr 2023

- Developed a cross-platform mobile application using React Native and Expo to showcase AI-generated fashion designs, integrating a pre-trained DCGAN model converted to TensorFlow Lite for on-device inference. Implemented a user-friendly interface allowing users to generate and browse fashion design concepts with 80% user satisfaction in usability testing.
- Built the mobile frontend with JavaScript/React Native, utilizing React Navigation for multi-screen navigation and Async Storage for local data persistence. Integrated Axios for API communication and implemented basic image handling with React Native Image Picker to allow users to upload reference images for style inspiration.
- Deployed a lightweight Flask REST API backend hosted on Heroku (free tier) to serve model predictions, with image preprocessing handled using Pillow and NumPy. Created the application UI using React Native Paper for Material Design components and tested the app on Android Studio Emulator and Expo Go for iOS/Android compatibility.

Technologies

- **Programming Languages:** Python, R, SQL, PySpark
- **ETL Tools:** AWS Glue, Python, PySpark, AWS EMR, Databricks
- **Cloud Technologies: AWS** – S3, Glue, EMR, RDS, Aurora, and Redshift
- **Database:** Oracle, SQL Server, MySQL, PostgreSQL, Redshift
- **BI Tools:** AWS QuickSight, Tableau, PowerBI, Excel
- **AI-ML Tools:** AWS Bedrock, Amazon Q
- **Frontend:** HTML, CSS3, JavaScript, Bootstrap, Responsive Web Design
- **Backend:** Python Flask, JSON
- **Deployment and Hosting:** AWS EC2, GitHub, Git, MongoDB
- **Model Integration:** TensorFlow, Keras, REST APIs
- **IDEs:** Jupyter Notebook, PyCharm, VS Code
- **Data Packages:** NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, TensorFlow

Certifications

- **AWS Certified Data Engineer – Associate**
  Issued by: **AWS**
- **AWS Certified Cloud Practitioner**
  Issued by: **AWS**
- **Smart India Hackathon-2022**
  Issued by: Institution's Innovation Council