

Sai Anuragh Sangoju

Dallas, Texas

+1-945-274-8276 sangojusaianuragh@gmail.com [LinkedIn](#) [Github](#)

SUMMARY

Data Scientist with 1.5 years' experience in building and optimizing data pipelines and predictive models. Skilled in Python, machine learning, and advanced data analytics to enhance forecasting and operational efficiency. Demonstrated success in developing dynamic dashboards and ETL processes, complemented by ongoing Master's studies in Business Analytics and AI.

EDUCATION

University of Texas at Dallas

Master of Science, Business Analytics and Artificial Intelligence

Aug 2023 - May 2025

B V Raju Institute of Technology

Bachelor of Technology

Jul 2018 - Jun 2022

EXPERIENCE

Cognizant Technology Solutions | *Programmer Analyst*

Feb 2022 - Jun 2023

Optimized ETL pipelines using Informatica PowerCenter to extract, transform, and load approximately 5TB monthly of large datasets into enterprise data warehouses, enhancing data processing efficiency.

- Built regression and classification models using Python, scikit-learn, and TensorFlow to achieve 85% accuracy while integrating advanced machine learning techniques into data science solutions.
- Engineered a predictive model with Python and TensorFlow that increased forecast accuracy by 25% for client projects, contributing to improved investment research insights.
- Developed dynamic Tableau dashboards with real-time data integration to support stakeholder decision-making and reduce manual analysis efforts.
- Conducted exploratory data analysis using SQL and Python to identify key business metrics, resulting in a 10% improvement in operational efficiency.
- Performed data cleaning and preprocessing on large datasets using Pandas and NumPy, reducing data inconsistencies by 25% and ensuring reliability for both structured and unstructured data analyses.
- Monitored and optimized pipeline performance to achieve a 30% reduction in processing time, playing a key role in the development of scalable data pipelines.

GenC Next | *Intern*

Contributed to ETL processes and data preprocessing tasks using Pandas, ensuring data quality and readiness for further analytical modeling.

PROJECTS

Optimizing regional strategies and Product Offerings for Gardein in the Meat Substitute Market

- Identified the Northeast and California as the top-performing regions, generating \$40.5M and \$31.3M in sales, respectively, for flagship products like "Burger Patty" and "Ground Regular".
- Observed a 30% sales increase during winter for flavors like "Classic" and "Regular", while summer sales lagged by 20%.
- Proposed strategies to boost sales of underperforming products, such as phasing out low demand items (e.g., Boxed "Burger" Regular with <200k sales) and enhancing visibility for high-potential products like "Burger patty" in resealable bags.

Conversational AI Chatbot for automated Food Ordering

- Developed NLP based intent recognition using training phases and rule-based logic in FastAPI, handling 10+ custom intents for food orders, modifications, and tracking.
- Optimized chatbot responses by structuring training phases in a text corpus, achieving 90% accuracy in intent recognition.
- Developed REST API end points in FastAPI, ensuring easy integration with front-end applications or third-party services.

Machine Learning-Based Rainfall Classification

- Applied multiple classification algorithms including Linear Regression, K-Nearest Neighbors (KNN), Decision Trees, Logistic Regression and Support Vectors Machines (SVM) to identify patterns and predict outcomes.

SKILLS, ACTIVITIES & INTERESTS

- **Technical Skills:** NumPy, Pandas, SciPy, Python, Machine Learning, Deep Learning, Natural Language Processing, Large Language Models, Prompt Engineering, RAG, GenAI, Statistical Modeling, TensorFlow, PyTorch, Scikit-learn, Linux, SQL, PySpark, API Development, FastAPI, Docker, Git
- **Certifications & Training:** Graduate Certificate in Applied Machine Learning, Tableau 2022 A-Z: Hands-On Tableau Training for Data Science