

Teja Sri V



United States | vtejasri15@gmail.com | 913-267-9839 | yourwebsite.com | LinkedIn | GitHub

Summary

Results-oriented Data Engineer passionate about helping financial institutions turn complex, fast-moving data into clear, strategic advantage. Proven ability to support risk mitigation, transaction monitoring, and audit-ready pipelines through automation and efficient data orchestration.

Education

University of Central Missouri, MS, Computer Science —MO,USA | GPA: 3.6/4.0 | Aug 2023 – May 2025
Coursework: Big Data and Analytics, Statistical Foundation and AI, Machine Learning, Cloud Computing

Gudlavalleru Engineering College, Bachelors, Computer Science — India | GPA: 8.5/10 | Aug 2018 – May 2022
Coursework: Advanced Algorithms, Computer Networks, NoSQL, Cybersecurity, Python

Technical Skills

- **Programming:** Python (Pandas, NumPy, SciPy, Scikit-learn, Matplotlib), R, SQL, Java, Javascript, Html, CSS
- **Data Visualization:** Tableau, Power BI, Matplotlib, Seaborn
- **Big Data Technologies:** PySpark, Apache Spark, Databricks, Snowflake, Datalake, Airflow
- **Cloud Platforms:** AWS (Redshift, Sagemaker), Azure (Databricks, Synapse, Data Factory)
- **Statistical Analysis & Machine Learning:** Regression Analysis, Forecasting, A/B Testing, Hypothesis Testing

Experience

Data Engineer Intern — Edupanda Advanced Solutions, Telangana, India Jan 2023 – Jul 2023

- Built and automated ETL pipelines using PySpark on Databricks to process retail sales data
- Designed 20+ Airflow DAGs for orchestrating batch and streaming workflows, increasing pipeline reliability
- Implemented schema validation and quality checks for incoming datasets, boosting accuracy by 63% ; enabled Power BI integration.
- Tuned AWS services (S3, EC2, Glue) for scalable transformation, lowering costs, and enhancing throughput.

Python Developer Intern — Icompaas, Hyderabad, India Aug 2022 – Dec 2022

- Developed microservices using Flask and FastAPI to build internal data APIs, enabling integration of structured datasets across distributed systems.
- Reduced system overhead by 20% through modular React.js components and Redux, improving API performance monitoring and frontend responsiveness.
- Streamlined build and release dataflow via CI/CD pipelines using GitHub Actions, resulting in a 70% reduction in deployment failures and improved overall system stability.

Projects

Distributed NLP Data Pipeline for Multilingual Review Analysis [GitHub] Oct 2024

- Architected a multilingual customer review analysis system using distributed NLP and LDA to extract trends.
- Containerized processing stages for reproducibility and applied memory slicing to improve throughput by 22%.
- Introduced topic drift tracking and schema-flexible input handling, enabling real-time feedback insights that simulated a 12% boost in customer retention.

Retail Pricing Strategy System with Data Orchestration and Machine Learning [GitHub] Nov 2024

- Built a real-time pricing engine using Spark SQL and RLLib, increasing revenue by 13% and reducing stockouts by 18% through dynamic pricing aligned with inventory flow.
- Deployed modular pipelines on AWS using a workflow scheduler and Docker, enhancing automation by 25% and enabling scalable multi-region deployment.