

GOUTHAMI NADUPURI

San Francisco Bay Area, California | +1 (341) 356-3732 | gouthaminadupuri25@gmail.com | [LinkedIn](#) | [GitHub](#)

PROFESSIONAL SUMMARY

Software Engineer with 4+ years of experience building scalable data pipelines, distributed processing systems, and production-grade machine learning solutions. Strong foundation in Python, Java, Spark, and cloud-native systems with hands-on experience designing Retrieval-Augmented Generation (RAG) architectures, NLP pipelines, and large-scale analytics systems. Passionate about building AI-powered systems that improve knowledge discovery and decision-making. Actively seeking opportunities in the San Francisco Bay Area, California.

TECHNICAL SKILLS

Languages: Python, Java, SQL, JavaScript, Typescript

Backend & Systems: REST APIs, Microservices, Spring Boot, Distributed Systems

ML & NLP: Scikit-learn, TensorFlow, Embeddings, Semantic Search, RAG

Data & Infra: Apache Spark, PySpark, ETL Pipelines, Data Modeling

Databases: PostgreSQL, Vector Databases

Cloud & DevOps: AWS, GCP, Docker, Git, CI/CD

EDUCATION

University of the Pacific, San Francisco, CA — M.S. Data Science (Aug 2023 – May 2025)

GITAM University, Visakhapatnam, India — B.Tech. Computer Science & Engineering (Jul 2008 – May 2012)

EXPERIENCE

Data Scientist (Part-Time) — Symosis Security (Sunnyvale, US) | 2025

- Built cybersecurity analytics dashboards using Tableau, Power BI, and Databricks to support reporting and decision-making.
- Engineered data transformation workflows in Databricks to clean, enrich, and standardize structured datasets.
- Developed reusable analytics pipelines in collaboration with Data Engineering and Cybersecurity teams.
- Contributed to sprint reviews and cross-functional discussions to improve reporting efficiency and delivery.

Data Quality Analyst — Vitech Systems Group (Hyderabad, India) | 2020–2023

- Built scalable Spark-based ETL pipelines processing large financial datasets across distributed systems.
- Automated anomaly detection and validation systems improving reporting accuracy by 45%.
- Designed A/B testing frameworks to evaluate system and data pipeline improvements.

Senior Analyst (Google Client) — Wipro (Hyderabad, India) | 2018–2019

- Applied SQL and statistical modeling to detect fraud patterns in large-scale e-commerce datasets.
- Built risk analytics dashboards improving detection precision and reducing false positives.

Process Associate — Accenture (Hyderabad, India) | 2016–2017

- Analyzed 1M+ transactional records using SQL and Java, improving reconciliation accuracy by 15%.
- Executed large-scale data validation workflows to improve reporting reliability.

PROJECTS

AI-Powered Enterprise Search (RAG System)

- Built end-to-end RAG architecture with document ingestion, embedding generation, vector indexing, semantic retrieval, and LLM inference.
- Improved retrieval relevance through chunking optimization and evaluation pipelines to reduce hallucinations.

NLP Classification Pipeline

- Developed text classification system using TF-IDF and ML models achieving 85–90% accuracy with precision/recall/F1 evaluation.

Time-Series Forecasting System

- Built forecasting models (ARIMA, XGBoost, LSTM) with walk-forward validation achieving 20–35% MAPE improvement.

CREDENTIALS & VOLUNTEERING

- GDG Organizer- Organized developer events, coordinated community activities, and developer workshops.
- AWS Cloud Practitioner Essentials
- Google Cloud Certification: Machine Learning Engineer
- IBM RAG and Agentic AI