GOUTHAMI NADUPURI

+1(341)-356-3732 | gouthaminadupuri25@gmail.com | LinkedIn | GitHub

TECHNICAL SKILLS

Programming: Python, C++, SQL, JavaScript, Data Structures, Algorithms, Software Development Lifecycle (SDLC)

Machine Learning & Artificial Intelligence: Supervised and Unsupervised Learning, Deep Learning (TensorFlow, PyTorch, Keras), NLP, LLMs, RAG, Predictive Modeling, Model Evaluation (RMSE, MAPE, F1), Feature Engineering

Data Engineering & Cloud: ETL Development, Apache Spark, Apache Kafka, AWS, dbt, Data Modeling (Star Schema, SCD, Temporal), Data Pipelines, CI/CD, Docker, Kubernetes, BigQuery, Lakehouse

Tools & Platforms: Jupyter, Google Colab, VS Code, PyCharm, Git, Jira, Confluence, Microsoft Project, Postman, Docker, Kubernetes

Data Governance & Documentation: Data Quality, Lineage, GDPR/CCPA Awareness, Access Management, Technical Documentation

Data Analytics & Visualization: Matplotlib, Seaborn, Plotly, Tableau, Power BI, Looker

Statistical Analysis & Scientific Computing: Hypothesis Testing, Regression Analysis, A/B Testing, Bayesian Inference, Probability & Statistics, Linear Algebra, NumPy, SciPy, Pandas

EXPERIENCE

Data Quality Analyst (Jan 2020 – June 2023)
Vitech Systems Group - Hyderabad, India
- Project: MPI & NBTF
- Extracted and validated large datasets related to pension and insurance plans for analysis and reporting.
- Conducted data quality checks, statistical analysis, and root cause investigations to ensure accuracy.
- Developed SQL queries and Excel models to analyze member data, track claims, and generate reports.
- Built dashboards to provide real-time visibility into claims processing and member benefits.
- Conducted data quality analysis on pension and healthcare datasets to identify anomalies and compliance issues.
- Automated data validation and reporting tasks using SQL and Python scripts.
- Created dashboards and visualizations to present insights to client stakeholders.
- Partnered with business users to convert requirements into analytical reports.
- Recognized for improving accuracy and reliability of business-critical data.

Senior Analyst (Mar 2018 – Nov 2019)
Wipro - Hyderabad, India
- Project: Google Trust & Safety
- Conducted in-depth data analysis on e-commerce merchant activities to ensure compliance with Google Shopping policies.
- Designed and maintained data quality reports, identifying violations and providing actionable insights.
- Automated recurring reporting workflows, reducing manual effort by streamlining data processes.
- Trained junior analysts in data interpretation, visualization, and reporting best practices.
- Delivered insights to improve customer trust and user experience.

Data Associate (Aug 2016 – Dec 2017)
Accenture - Hyderabad, India
- Project: Geospatial
- Processed and analyzed large-scale geospatial and mobility datasets from Uber Maps to improve route accuracy and ETA prediction.
- Developed and optimized SQL and Python scripts for data extraction, cleaning, and transformation of GPS, trip, and traffic data.
- Built interactive data visualization to display traffic density, route efficiency and performance metrics.

EDUCATION

University of the Pacific   Aug 2023 – May 2025
Master's of Science in Data Science, San Francisco, CA
Gitam University   Jul 2008 – May 2012
Bachelor of Technology in Computer Science and Engineering, Vizag, India

PROJECTS

Customer Churn Prediction | Machine Learning

- Delivered retention strategy insights based on customer service usage and behavioral patterns, improving churn prediction accuracy.
- Created visualizations (Matplotlib, Seaborn) to communicate insights to non-technical stakeholders.

Fake News Detection | Natural Language Processing (NLP)

- Built an NLP pipeline (text preprocessing, tokenization, TF-IDF, word embeddings) for news classification.
- Implemented Logistic Regression, Random Forest, and BERT-tiny models for performance benchmarking.
- Achieved high accuracy and F1-score, with BERT-tiny outperforming classical models.
- Visualized model comparisons and misclassifications using Seaborn & Matplotlib.

E-Commerce Fraud Detection System | Data Engineering

- Designed a real-time fraud detection pipeline using Apache Kafka for streaming transaction data.
- Integrated machine learning models for anomaly detection, risk scoring, and automated fraud alerting.
- Used Elasticsearch + Kibana dashboards for anomaly visualization and fraud monitoring.
- Implemented ETL workflows for ingesting, cleaning, and transforming high-volume transaction data.

Time Series Forecasting | Data Science

- Built ARIMA and Prophet models to forecast sales with trend and seasonality decomposition.
- Conducted exploratory data analysis (EDA) and feature engineering to identify seasonality and outliers.
- Evaluated model performance using RMSE and MAPE, achieving improved forecasting accuracy.
- Developed visualizations and reports to support business planning and decision-making.

CREDENTIALS & VOLUNTEER CONTRIBUTIONS

- AWS Cloud Practitioner Essentials
- Python Programming Masterclass with Data Structures and Algorithms
- IBM RAG and Agentic AI
- Volunteer and active team member in Google Developers Group (GDG) community, supporting event organization, knowledge-sharing sessions, and community engagement initiatives.