Gouthami Nadupuri
Fremont, California | gouthaminadupuri25@gmail.com | +1-341-356-3732
LinkedIn | GitHub

PROFESSIONAL SUMMARY

Data Scientist with expertise in Python, R, SQL, Machine Learning, and AI, skilled in building predictive models and automating workflows for actionable insights. Strong foundation in data-driven decision-making to solve real-world problems. Currently pursuing a Master's in Data Science and seeking opportunities to apply analytical expertise in dynamic environments.

EDUCATION

- M.S. in Data Science – University of the Pacific, San Francisco, CA (2023 – 2025)
- B.Tech. in Computer Science – GITAM University, India (2008 – 2012)

EXPERIENCE

Data Quality Analyst Jan 2020 – Jun 2022
Vitech Systems Group Hyderabad, India

- Automated data validation and model testing using Python, SQL, and Postman to enhance efficiency and accuracy.
- Developed predictive models for pension & insurance clients, improving compliance accuracy by 35% and reducing regulatory violations by 20%.
- Designed ETL workflows for data ingestion, increasing data processing speed by 45%.
- Led data quality monitoring, ensuring adherence to Agile methodologies and regulatory standards.

Senior Data Analyst Jan 2018 – Nov 2019
Wipro Technologies Hyderabad, India

- Conducted exploratory data analysis (EDA) and developed BI dashboards in Tableau & Power BI to enhance data-driven decision-making.
- Automated data pipelines using Python and SQL, reducing ETL execution time by 90% and improving data accuracy by 40%.
- Applied A/B testing and KPI analysis to optimize financial services operations.
- Mentored and trained junior analysts, contributing to a increase in team efficiency.

Data Analyst Aug 2016 – Dec 2017
Accenture Solutions Pvt Ltd Hyderabad, India

- Developed customer segmentation models to optimize targeted marketing strategies.
- Built a real-time fraud detection system, reducing fraudulent transactions by 95%, saving millions in potential financial losses.
- Designed ETL pipelines to streamline data processing and improve accessibility.

SKILLS

- **Software Languages:** Python, R, Java
- **Database systems:** SQL, Apache Kafka, AWS (EC2, S3), ETL pipelines, NoSQL
- **Machine Learning:** TensorFlow, Time Series Forecasting, NLP, Linear/Logistic Regression, Decision Trees, Random Forests, Naive Bayes, k-NN, XGBoost, Gradient Boosting (Cat Boost, LightGBM), Support Vector Machines (SVM), Neural Networks, K-Means, Hierarchical Clustering, Principal Component Analysis (PCA), Singular Value Decomposition (SVD)
- **Data Visualization:** Tableau, Matplotlib, Seaborn, Power BI, Plotly
- **Tools & Platforms:** Jupyter Notebooks, Git, Postman, Google Collaboratory, Visual Code, PyCharm

PROJECTS

Analysis of U.S. Consumer Mortgage Complaints | Data Analysis

- Conducted exploratory analysis on consumer complaints to identify patterns and improve resolution strategies.
- Applied regression analysis, decreasing mortgage response time by 25%, improving customer satisfaction scores.

E-Commerce Fraud Detection System | Data Engineering

- Developed a real-time fraud detection system, leveraging Apache Kafka for data streaming, cutting fraud incidents by 40%.
- Integrated ML models with ETL pipelines, achieving 95% fraud detection accuracy and reducing chargeback losses by 30%.
- Created interactive dashboards for anomaly detection and transaction monitoring using Kibana.

Fake News Detection | NLP

- Built an NLP-based pipeline to classify news articles, achieving 90% accuracy in identifying misinformation.
- Implemented Logistic Regression, Random Forest, and BERT-tiny, boosting classification performance by 25%.
- Utilized Python libraries to visualize classification metrics and enhance interpretability.

Heart Disease Prediction | Machine Learning

- Developed a predictive model for heart disease detection, achieving 85% accuracy, assisting in early diagnosis.
- Implemented 5 classification algorithms (Logistic Regression, Random Forest, SVM, Gradient Boosting, KNN), boosting model efficiency by 20%.
- Evaluated model performance using ROC-AUC and confusion matrix, reducing false positives by 30%.