Hima Bindu Gayathri | Data Scientist

Denton, TX | +1 (469) 260-0258 | gayathri260501@gmail.com | www.linkedin.com/in/hima-bindu-gayathri-134a31208

SUMMARY

Data Scientist with 3+ years of experience driving business value through data modeling, statistical analysis, and machine learning. Proven track record of transforming complex data into actionable insights to support decision-making and strategy. Skilled in Python, SQL, and cloud platforms like AWS. Adept at collaborating across teams to deploy scalable data solutions and predictive models in production environments.

SKILLS

Machine Learning & AI: Supervised Learning (Linear/Logistic Regression), Unsupervised Learning (Clustering, K-Means), Neural Networks & Deep Learning (CNNs, RNNs, GANs), Reinforcement Learning, NLP, Sentiment Analysis, Text Classification, Model Evaluation & Tuning
Programming Languages: Python (NumPy, Pandas, Scikit-learn, TensorFlow, Keras, PyTorch), R, SQL, SAS, Scala, Luigi
Data Analysis & Statistics: Statistical Modeling, A/B Testing, Time Series Forecasting, Exploratory Data Analysis (EDA), Data Cleaning & Wrangling, Data Preprocessing, DataBricks
Big Data Technologies: Hadoop, Apache Spark, Apache Kafka, Hive, Apache Flink, Apache NiFi
Data Visualization: Tableau, Power BI, D3.js, Plotly, Seaborn, Matplotlib
Databases: SQL (MySQL, PostgreSQL, Oracle), NoSQL (MongoDB, Cassandra, Redis, HBase), Data Warehousing (Amazon Redshift, Snowflake)
Cloud Platforms: Amazon Web Services (AWS), Microsoft Azure, AWS SageMaker, Lambda, EC2, S3, Google AI Platform
Soft Skills: Problem Solving, Project Management, Agile (Scrum/Kanban), Business Intelligence & Decision Making, Best Practices

CERTIFICATIONS

Microsoft Certified: Azure Data Fundamentals
Microsoft Certified: Azure Fundamentals
Salesforce Certified: Tableau Desktop Foundations
Databricks Certified: Generative AI Fundamentals

EXPERIENCE

Principal Financial, Data Scientist **Dec 2024 – Present | TX**

- Developed an XGBoost-based credit risk model for consumer lending, improving default prediction accuracy by 19% and reducing portfolio-level loss estimates by $2.4M annually after full deployment.
- Created a scalable time-series forecasting solution using Prophet for insurance claim volumes, reducing forecast deviation from actuals by 14% and helping actuaries adjust monthly reserve estimates more accurately.
- Automated quarterly regulatory reports using Apache NiFi, SQL, and Python scripts, reducing manual data processing time by 38 hours per quarter and minimizing risk of compliance errors.
- Built customer segmentation models using K-Means on 3 million+ policyholder profiles, enabling targeted product recommendations that increased cross-sell conversion rates by 11% within three months of rollout.
- Integrated AWS Lambda with EC2-hosted scoring APIs to enable near-real-time credit scoring for new applications, reducing processing latency by 2.3 seconds per transaction.
- Conducted bias audits using SHAP values and fairness metrics, identifying age-related bias in two models and reducing disparity in prediction outputs by 6% after model retraining.

Mphasis, Data Scientist **Jun 2022 – Dec 2023 | India**

- Built BERT-based NLP models for automated contract clause classification, reducing manual document review time by 41% and increasing extraction accuracy on edge cases by 18% across 20,000+ documents.
- Designed monitoring dashboards using Power BI to track live model metrics for 15+ client projects, reducing SLA violations by 27% and enabling proactive alerts for data drift or pipeline errors.
- Developed a Python and SQL-based A/B testing framework that supported over 200,000 daily users and allowed marketing teams to evaluate conversion changes with 95% confidence in under 48 hours.
- Processed real-time financial transactions using Spark Structured Streaming and Apache Kafka, reducing anomaly detection latency by 3.2 seconds and supporting immediate transaction flagging for review.
- Collaborated with cloud teams to deploy containerized models on Azure Kubernetes Service, improving CI/CD integration and reducing deployment turnaround time by 42% across multiple clients.
- Performed root cause analysis for model underperformance in quarterly audits, uncovering schema drift issues that impacted 3 production models and implementing fixes that restored baseline accuracy within days.

Ascent SoftTech, Data Analyst **Jun 2021 – May 2022 | India**

- Analyzed transactional and behavioral data from five SME clients to identify workflow bottlenecks, leading to a 9% reduction in task processing times after targeted changes to internal tooling processes.
- Designed monthly Tableau dashboards for C-level reporting on KPIs like churn, CAC, and LTV, decreasing manual reporting time by 80% and improving internal visibility for decision-makers.
- Built complex SQL queries integrating relational and NoSQL data sources across PostgreSQL, MongoDB, and Redis to support seven key weekly reports for sales, operations, and product teams.
- Standardized inconsistent metadata across 100,000+ product SKUs using Python scripts, improving search relevance and product categorization in internal tools and client-facing portals.
- Created DAX-powered Power BI dashboards to monitor internal helpdesk ticket resolution times, which helped reduce open ticket aging beyond SLA by 36% within two quarters.
- Implemented time-series analysis on historical transaction data to forecast short-term demand for two clients, reducing stockout incidents by 17% across key SKUs during pilot implementation.

EDUCATION

Master in Information Systems and Technologies **Jan 2024 – May 2025**
University of North Texas | Denton, Texas

Bachelor in Electrical and Electronics Engineering **Aug 2018 – May 2022**
R.V.R & J.C College of Engineering | Andhra Pradesh, India

PROJECTS

Financial Risk Forecasting with Machine Learning

- Built predictive models using logistic regression and random forests, increasing forecast accuracy of financial risk scenarios by 25% across high-value portfolios.
- Evaluated model accuracy, precision, and recall using cross-validation and confusion matrix analysis in Python (scikit-learn).

Accelerated ETL with Airflow and SQL

- Designed and scheduled modular DAGs in Apache Airflow, reducing ETL latency by 40% for daily financial reporting jobs.
- Streamlined complex SQL transformations across raw, staging, and analytics layers, improving end-to-end data pipeline reliability.