

Sathwik Reddy Velmala

Data Scientist

sathwikr@jobhuntmails.com | (737) 444-6526 | Houston, TX – 77025 | [LinkedIn](#) | [GitHub](#) | [Portfolio](#)

Summary

Results-driven Data Scientist with 4+ years of experience developing scalable machine learning models, robust data pipelines, and actionable insights across domains, including fraud detection and churn prediction. Proficient in developing ETL workflows integrating data from diverse sources, analyzing large-scale behavioral data, performing advanced feature engineering, and driving segmentation for personalized targeting. Skilled in deploying ML models on cloud and implementing Git-based CI/CD workflows for automated releases. Experienced in statistical testing to assess model performance and designing interactive dashboards in Tableau and Power BI to visualize trends. Proven ability to collaborate cross-functionally and communicate complex insights to drive business decisions.

Technical Skills

- **Programming Languages:** Python, SQL, Java
- **Machine Learning & AI:** Predictive Modeling, XGBoost, LightGBM, Random Forest, Decision Trees, Time-Series Forecasting, Clustering, NLP, Transformer Models, Generative AI, LLMs, TensorFlow, PyTorch, scikit-learn
- **Statistical Analysis:** Feature Engineering, Customer Segmentation, EDA, Advanced Excel, A/B Testing, Multivariate Testing
- **Data Engineering & Processing:** Databricks, PySpark, Apache Spark, NumPy, Pandas, ETL Pipelines, Apache Airflow
- **Data Visualization & Reporting:** Tableau, Power BI, AWS QuickSight, Matplotlib, Seaborn, KPI dashboards
- **Databases & Data Warehousing:** MySQL, Amazon Redshift, Snowflake, MongoDB, SQL Server, MariaDB
- **Cloud Platforms:** AWS (S3, SageMaker, Lambda, Glue, Athena), Azure (Azure ML, Blob Storage, Synapse, Azure Data Factory, Azure DevOps, Azure OpenAI, Document Intelligence)
- **DevOps & Others:** Git, Docker, Jenkins, Version Control, GitHub Actions, CI/CD Pipelines, Microsoft Office, FastAPI, Postman

Professional Experience

Data Scientist, Red Ventures

01/2024 – Present | Remote, USA

- Built and deployed predictive models using **Python (XGBoost, LightGBM, TensorFlow)** on **AWS SageMaker** to forecast churn and CLV, informing personalized retention strategies and reducing customer churn by 12%.
- Developed and automated ETL pipelines using **AWS Glue** and **Apache Airflow** to clean and transform customer and transactional data from **Snowflake, MongoDB**, and **S3** into **Redshift**, reducing data latency by 25% and enabling analytics-ready data access.
- Performed data analysis and feature engineering on behavioral and time-series data using **Python, SQL**, and **PySpark** to build segmentation models, providing personalized recommendations and boosting customer engagement by 20%.
- Applied statistical modeling techniques such as **A/B Testing, Multivariate Testing**, and **Uplift Modeling**, to assess model performance and evaluate marketing campaign strategies, resulting in a 20% improvement in ROI.
- Designed interactive dashboards in **Tableau** to monitor churn trends, customer engagement, and retention KPIs, cutting manual reporting time by 40% and accelerating data-driven decision-making for business teams.
- Collaborated with product and engineering teams in **Agile** settings to define KPIs, communicate model outcomes, and assist in deploying **CI/CD MLOps** workflows with **Git** and **Jenkins**, cutting manual intervention by 50%.

Data Scientist, Genpact

02/2021 – 12/2022 | India

- Built and automated ETL pipelines using **AWS Glue** and **Apache Airflow** to integrate transactional, user, and behavioral data from **Amazon S3, SQL Server**, and **Amazon Redshift**, increasing ETL processing efficiency by 30% and reducing overall data latency.
- Developed fraud detection models using **Python, XGBoost**, and **Random Forest** to classify high-risk transactions and optimize fraud detection workflows, achieving 95% model accuracy and contributing to a 20% increase in detection accuracy.
- Trained and deployed real-time fraud detection models on **AWS SageMaker** with automated hyperparameter tuning and model evaluation techniques, and integrated with **AWS Lambda** for low-latency inference, enabling high-accuracy inference workflows.
- Engineered fraud-specific features from user and transaction data and performed data analysis using **Python, SQL**, and **Excel** to uncover anomalous patterns, improving model precision by 15% and reducing false positives by 25%.
- Created real-time monitoring dashboards in **Power BI** and **AWS QuickSight** to visualize fraud trends, track KPIs, and provide actionable insights to the fraud prevention team, reducing manual reporting time by 30% and improving decision-making.

Data Analyst, Gitcore Solutions PVT LTD

12/2019 – 07/2020 | India

- Analyzed user behavior data using **Excel** to inform product feature prioritization, boosting adoption of prioritized features by 15%.
- Performed data validation and cleansing to identify anomalies and missing values, enhancing product data accuracy by over 20%.
- Migrated complex product usage data from **SQL Server** to **MariaDB**, ensuring minimal downtime through schema normalization.
- Designed interactive dashboards in **Power BI** to track feature adoption, user drop-off rates, and KPI trends, helping stakeholders quickly assess product performance and make data-driven decisions, reducing drop-off rates by 18% over 3 quarters.

Education

Master of Science, University of Houston

01/2023 - 05/2024 | Houston, TX, USA

Engineering Data Science

Bachelor of Technology, Malla Reddy Engineering College

08/2017 - 06/2021 | Hyderabad, India

Computer Science and Engineering