Sathwik Reddy Velmala Data Scientist

sathwikr@jobhuntmails.com | (737) 444-6526 | USA | LinkedIn | GitHub | Portfolio

Summary

Results-driven Data Scientist with over 3+ years of experience in predictive modeling, fraud detection, and real-time data analysis within e-commerce and retail domains. Proficient in Python, SQL, and advanced ML frameworks like XGBoost, TensorFlow, and PyTorch. Expertise in deploying scalable models using cloud platforms such as AWS (SageMaker, Lambda, Glue), Azure, and GCP. Skilled in big data tools (PySpark, Apache Kafka), building efficient ETL pipelines, and delivering actionable business insights through data visualization tools like Tableau and Power BI. Adept at solving complex data challenges and optimizing model performance in dynamic environments.

Technical Skills

- **Programming Languages:** Python, SQL, R
- **Machine Learning & AI:** XGBoost, Random Forest, Isolation Forest, Autoencoders, CNN, RNN, LSTM, GANs, ARIMA, SVM, Decision Trees, K-Means Clustering, Transformer Models, Scikit-learn, TensorFlow, PyTorch, LightGBM, SHAP, LIME
- **Cloud Platforms:** AWS (SageMaker, Lambda, Glue, Redshift, Kinesis), Azure (Machine Learning, Data Lake, Blob Storage), GCP (BigQuery, Google Cloud Storage, AI Platform)
- **Big Data & Data Engineering:** Apache Spark, PySpark, Hadoop, Hive, Apache Kafka, ETL Pipelines, Apache Airflow
- **Data Visualization & Reporting:** Tableau, Power BI, AWS QuickSight, Matplotlib, Seaborn
- **Databases:** MySQL, PostgreSQL, MongoDB, Amazon Redshift, Snowflake, MS SQL Server, NoSQL
- **Other Tools & Technologies:** Docker, Kubernetes, Jenkins, JIRA, ServiceNow, Pandas, NumPy, Git, AWS CloudWatch

Professional Experience

Data Scientist, *Red Ventures* 01/2024 – Present | Remote, USA

- Designed and implemented automated ETL pipelines using Apache Airflow and AWS Glue, integrating data from diverse sources such as Amazon S3, Snowflake, and MongoDB for real-time churn prediction.
- Leveraged Databricks and PySpark to process large datasets, extracting key features from customer behavior, transaction history, and demographics, significantly improving model accuracy and performance.
- Built and optimized churn prediction models using advanced machine learning algorithms like XGBoost, LightGBM, and TensorFlow, achieving 94% accuracy and reducing churn by 25%.
- Deployed models to AWS SageMaker, utilizing AWS Lambda for real-time inference and automated hyperparameter tuning to scale churn prediction across millions of customer records. Integrated with CRM systems for targeted retention actions.
- Developed interactive dashboards in Power BI, enabling stakeholders to monitor churn trends, identify high-risk segments, and evaluate the effectiveness of retention strategies in real-time.
- Incorporated Explainable AI (XAI) tools like SHAP and LIME to provide transparent model explanations, ensuring stakeholder trust and actionable insights to inform customer engagement and retention campaigns.

Data Scientist, *Genpact* 02/2021 – 12/2022 | India

- Developed fraud detection models using XGBoost and Isolation Forest, achieving 95% accuracy in identifying fraudulent transactions, leveraging Python and Scikit-learn to implement machine learning algorithms for scalable fraud prevention.
- Trained and deployed fraud detection models on AWS SageMaker, using automated hyperparameter tuning and model evaluation techniques to optimize performance, for real-time, high-accuracy fraud detection in dynamic e-commerce environments.
- Integrated AWS Kinesis for real-time transaction data streaming, enabling immediate fraud detection and reducing transaction processing time by 40%. Used AWS Lambda to deploy models with minimal latency for real-time decision-making.
- Built and optimized ETL pipelines with AWS Glue to clean, transform, and load high-velocity transaction data from Amazon S3 into machine learning models, streamlining data processing and ensuring accurate predictions.
- Utilized PyOD (Python Outlier Detection) library for anomaly detection, improving model precision by reducing false positives and enhancing overall fraud detection performance.
- Created real-time monitoring dashboards in AWS QuickSight to visualize fraud trends, track KPIs, and provide actionable insights to the fraud prevention team, improving decision-making and accelerating response times to emerging threats.

Data Analyst, *Gitcore Solutions PVT LTD* *12*/2019 – 07/2020 | India

- Conducted advanced data analysis using Python, Power BI, and Excel to uncover trends, enhancing decision-making and improving process efficiency by 15%.
- Ensured data integrity by identifying errors, missing values, and inconsistencies, increasing data accuracy by over 40%.
- Developed interactive dashboards and visualizations in Power BI, enabling stakeholders to easily interpret key performance metrics.
- Automated data processing tasks, reducing manual effort by 30% and ensuring timely delivery of actionable insights.

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