

Sravan Sai Methuku

Missouri, USA | +1(314)-312-1223 | sravansaimethuku45@gmail.com | GitHub | LinkedIn

SUMMARY

Data Scientist with 4+ years of experience in statistical modeling and predictive analytics to support data-driven decisions across healthcare, finance, and retail. Proven track record of transforming large datasets into actionable insights that enhance products and streamline processes. Skilled at communicating complex findings clearly to both technical and non-technical stakeholders.

TECHNICAL SKILLS

Programming Languages & Querying: Python (Pandas, NumPy, PySpark), SQL (PostgreSQL), R

Machine Learning & Statistical Analysis: XGBoost, Random Forest, Linear Regression, Logistic Regression, Hypothesis Testing, Bayesian Optimization, Time Series Forecasting, Causal Inference, A/B Testing, Predictive Modeling, Anomaly Detection, Price Elasticity Modeling, Fraud Detection

Model Development & Deep Learning: PyTorch, TensorFlow, Keras, Scikit-learn

Cloud & Data Engineering: AWS (S3, SageMaker, Glue), Azure (Databricks, Synapse, Data Factory), GCP (Vertex AI, BigQuery), Apache Spark, Hadoop, Kafka, Snowflake, ETL, Data Pipelines, Data Warehousing, Data Governance

MLOps & DevOps: MLflow, Airflow, Docker, Kubernetes, FastAPI, GitLab, CI/CD, Model Monitoring, Terraform, Evidently AI

Data Visualization & Reporting: Power BI, Tableau, Streamlit, Plotly, Dashboard Deployment, Business Analytics

Tools & Collaboration: Git, JIRA (Agile/Scrum), Data Storytelling, Stakeholder collaboration

Generative AI Tooling: Hugging Face Transformers, langChain (RAG, PEFT/lora), Pinecore, OpenAI API, Prompt Engineering

EXPERIENCE

Anheuser-Busch InBev

Aug 2024 – Present

Data Scientist

Saint Louis, MO

- Architected a multi-country price elasticity model to optimize strategic pricing, leveraging Python (XGBoost, scikit-learn) and PySpark on Databricks with 1.5M+ sales records and macroeconomic event data; achieved 90% accuracy and enabled automated pricing decisions across 5 markets.
- Designed and optimized end-to-end pricing workflows on Azure Synapse and Spark, performing feature engineering on 50M+ records and deploying interactive simulation tools via Power BI; improved pricing precision and shortened planning cycles by 40%.
- Developed LightGBM-based time series forecasting models in Snowflake SQL and Python to optimize production planning; improved forecast accuracy by 20% and reduced inventory shortages across 10 countries.
- Orchestrated automated model-retraining pipelines using Apache Airflow, engineering lag-based anomaly detection features that improved forecast stability and reduced manual intervention hours by 30%.
- Conducted A/B tests of dynamic pricing strategies for BEES B2B using CatBoost uplift modeling and TensorFlow Recommenders boosting basket size by 12% and incremental revenue by 20%.
- Mentored 4 junior data analysts, improving technical skills by 20% and enabling the team to deliver 3+ model insights that influenced C-level business decisions.

Centene Corporation

Dec 2023 – May 2024

Data Scientist / ML Engineer

Saint Louis, USA

- Architected Bayesian regression and SARIMA time-series disease-risk models in AWS SageMaker, analyzing longitudinal clinical data to forecast chronic conditions enhancing early-intervention capabilities and reducing projected treatment lag by 15
- Engineered unsupervised fraud detection pipeline using IsolationForest and PyTorch autoencoders on AWS Glue/EMR to analyze 50M+ insurance claims identified 17% anomalous patterns reducing false positives by 30% and prioritizing high-risk audits.
- Designed an NLP-based provider-note risktagging pipeline with spaCy and BERT, extracting structured anomalies from clinical notes improving investigation accuracy by 22% and reducing fraud case resolution time.

- Conducted data preprocessing and cleansing of 50M+ records using PySpark and Pandas, ensuring high-quality inputs for time series and regression models used in large-scale operational forecasting.
- Developed an LLM-based question-answering system using LangChain (RAG pipeline with OpenAI + Pinecone) to retrieve clinical protocols and documentation, reducing internal support response time by 40% and enhancing decision support for care teams.

Glenmark Pharmaceuticals

Sep 2022 – Aug 2023

Data Scientist

Bangalore, India

- Addressed inefficiencies in trial planning by building a Clinical Trial Outcome Prediction model using Python, SQL, Airflow, and XGBoost; conducted large-scale data analysis and PCA-based feature engineering, improving model performance and decision accuracy by 25%, leading to more informed go/no-go decisions
- Deployed XGBoost model via MLflow and Flask API; integrated predictions into Tableau for regulatory review, improving AUC-ROC to 0.87 and reducing trial design cycles by 20%, while enhancing data reliability and cross-functional collaboration in model-driven planning.
- Created Power BI and Tableau dashboards to monitor clinical KPIs and applied A/B testing and conversion analysis using Python, Pandas, and Scikit-learn; improved patient recruitment efficiency by 15% while supporting business strategy with data visualization.
- Developed biostatistical ML models using Random Forest, LightGBM, and TensorFlow with feature engineering, hyperparameter tuning, and SHAP; boosted accuracy by 30% and accelerated go/no-go decisions, integrating advanced predictive modeling techniques.
- Collaborated with clinical operations and production teams to integrate forecasting outputs into material planning processes, enabling on-time manufacturing and inventory alignment.

American Express

Mar 2021 – Aug 2022

Data Analyst

Bangalore, India

- Orchestrated the sourcing of real-time fraud data from 10+ internal and external sources, developing interactive dashboards with Power BI, Tableau, and Excel with DAX to enable business analytics. Built and implemented credit risk models using logistic regression, decision trees, XGBoost and H2O AutoML; applied SHAP and hypothesis testing to improve model precision by 22%, saving \$250K/quarter.
- Cleaned and prepared structured datasets using SQL and Pandas for classification models; enhanced feature reliability and improved model precision by 18% in operational risk scoring.
- Analyzed customer transactions and credit data to identify spending patterns, segment behavior, and detect anomalies using SQL, Python (pandas, seaborn), and Power BI, reducing churn by 18%.
- Improved data accessibility and Partnered with finance, compliance, and executive stakeholders to define KPIs and deliver insights that reduced risk exposure and supported quarterly revenue protection efforts.

EDUCATION

Southeast Missouri State University

Cape Girardeau, USA

Master of Science in Applied Computer Science, GPA: 3.7

May 2025

Relevant Coursework: Artificial Intelligence & Machine Learning Deep Learning & Computer Vision Natural Language Processing (NLP) Advanced Database Management Systems (SQL, NoSQL) Cloud Computing (AWS, Azure, GCP) Cybersecurity & Information Security Applied Data Science (Capstone Project) Data Mining & Feature Engineering Statistical Analysis & Hypothesis Testing MLOps & Model Deployment Big Data Processing with Spark Data Visualization & Reporting

CERTIFICATIONS

- AWS Certified Machine Learning - Associate.
- Databricks for Machine Learning