Suresh Kumar Maddala
Staff Data Scientist

Dallas, TX, USA | 248-949-3256 | msk88.in@gmail.com | https://www.linkedin.com/in/sureshkm-ai

PROFESSIONAL SUMMARY

Staff Data Scientist with 12+ years building forecasting and machine learning solutions across retail/CPG and supply chain analytics. Strong in **SQL, Python, and Databricks/Spark**, delivering production pipelines and model services at scale (e.g., **100K+ predictions/week**). Partner closely with business stakeholders to define metrics, forecast performance, and translate deep-dive analysis into executive-ready narratives. Experienced building self-serve dashboards/tools and applying GenAI selectively to accelerate insights and explain model outputs.

TECHNICAL SKILLS

Languages/Query: Python, SQL, R, PySpark
Data & Platforms: Azure Databricks (Spark/Delta), BigQuery, Vertex AI, AWS (S3/Glue/Athena/SageMaker), Azure ML
Forecasting & ML: Time-series forecasting (ARIMA, Holt-Winters, exponential smoothing), LightGBM/XGBoost, classification/regression, clustering, recommendation (SVD), NLP
MLOps/Production: MLflow, CI/CD (Azure DevOps), REST APIs, monitoring/anomaly detection, Docker/Kubernetes
BI & Storytelling: Power BI, Tableau, Looker, Streamlit, RShiny
GenAI: RAG, HuggingFace Transformers/Embeddings, LangChain/LangGraph
Experimentation & Stats: hypothesis testing, experimental design, statistical modeling

PROFESSIONAL EXPERIENCE

PepsiCo | Dallas, TX   10/2024 – Present
Staff Data Scientist

- Architected and deployed forecasting pipelines in **Azure Databricks (PySpark/Delta) + MLflow**, delivering weekly planning outputs at scale (**100K+ predictions/week**) with strong governance (versioning, reproducibility, monitoring).
- Built scenario-based forecasting (baseline vs. promotion) using **LightGBM + time-series methods**, improving forecast reliability and decision confidence for planning teams.
- Partnered with **Supply Chain and Finance** to define operational KPIs, align assumptions, and deliver leadership-ready narratives (drivers, sensitivities, risks) for S&OP and weekly business reviews.
- Implemented MLOps/CI/CD using **Azure DevOps + MLflow** with automated validation gates, rollout checks, and performance monitoring to reduce deployment cycle time and manual intervention.
- Led a **GenAI** Forecast Insights **Copilot** (**RAG + agentic** orchestration) with **Pinecone** indexing of forecast history, promo calendars, and metric definitions to deliver grounded Q&A and explanations.
- Implemented **guardrails** (structured outputs, retrieval-required prompts, standardized templates) to reduce **hallucinations** and improve explainability for business users.

Walmart | Dallas, TX   05/2022 – 10/2024
Senior Data Scientist

- Built a semantic "like-item" retrieval pipeline using **sentence-transformer embeddings** and **Pinecone vector search** to power buy-quantity optimization for cold-start/seasonal items, improving analog selection and business outcomes.

- Developed and productionized a **Club Alignment recommendation engine** using **SVD matrix factorization**, generating recommendations across **600+ clubs / 100+ departments** and delivering them through stakeholder-facing tools and APIs.
- Designed cloud-scale analytics + ML workflows on **BigQuery + Vertex AI**, improving inference latency and enabling secure, repeatable decision support for business teams.
- Built predictive models (XGBoost/LightGBM) for demand/basket signals with rigorous validation and operational adoption in downstream planning workflows.
- Developed an internal **agentic analytics assistant** (LLM + tool calling + RAG) that retrieved metric definitions and knowledge articles from **Chroma**, generated SQL with validation checks, and produced executive-ready summaries with traceability and safe-response behaviors.
- Optimized ML/**GenAI** services for reliability and efficiency (batching, caching, retry/timeout logic) and partnered with engineering to harden API interfaces for maintainability.

Niagara Bottling | Los Angeles, CA   04/2021 – 05/2022
Data Scientist
- Reduced annual transportation costs by **30%** by developing a spot market price prediction model (regression + gradient boosting), enabling **data-driven transportation pricing** and improved carrier selection decisions.
- Increased operational efficiency by **70%** by leading data engineering and optimization for a supply-chain planning initiative, reducing manual reporting and improving planning cycle time for operations teams.
- Delivered production-ready analytics/ML workflows (automated data pipelines, model scoring, and stakeholder-facing reporting) to support near real-time decision-making and scalable rollout.

Soothsayer Analytics | Detroit, MI   09/2019 - 03/2021
Data Scientist
- Built an LSTM-based forecasting solution that prevented production loss by **60%** through improved demand planning and inventory readiness.
- Designed automated anomaly detection using boosting techniques, reducing detection time from **4 days to 4 hours** for faster operational response.
- Implemented PySpark-based fraud modeling identifying duplicate claims with **94% recall**, reducing financial leakage.

Quadratic Insights | Hyderabad, IN   01/2014 - 08/2019
Data Scientist
- Improved fraud detection performance, **reducing false positives by 45%** by building and tuning **Random Forest / XGBoost** models; delivered monitoring views for anomaly trends and model drift.
- Built churn propensity models for financial clients, **reducing churn from 10% to 5%** and enabling retention strategies through targeted interventions and KPI tracking.
- Modernized data pipelines by migrating legacy **PL/SQL workflows to Spark/Hive-based ETL**, improving processing throughput by **4x** and automating orchestration via **Oozie** to meet SLA timelines.
- Developed demand forecasting models (**ARIMA, Holt-Winters**) to optimize inventory allocation, improving supply chain efficiency by **15%** and supporting planning decisions with clear forecast drivers.
- Implemented NLP-based automation for customer inquiries using **transformer-based models (BERT)** plus classical NLP, reducing average resolution time from **5 days to 2 days** and improving service responsiveness.
- Built customer segmentation using clustering (**K-means/Hierarchical/DBSCAN**), improving targeted marketing effectiveness and increasing engagement/conversions by **22%** through more personalized outreach.

EDUCATION
MTech, Artificial Intelligence — University of Hyderabad, India
BTech, Information Technology — JNTUK, India
Certifications: Databricks Certified Machine Learning (Associate/Professional)