AKSHAY AJITH JOSHI

Syracuse, New York • +1(315)-898-0031 • akjoshi@syr.edu • Linkedin • PORTFOLIO

SUMMARY

Data Scientist (MS Applied Data Science) with expertise in healthcare analytics, predictive modeling, and machine learning. Strong background in building scalable ETL pipelines, real-time forecasting models, and AI-driven solutions to optimize patient care, financial metrics, and hospital operations. I am skilled in AWS infrastructure, causal inference, NLP, and experimentation. Passionate about leveraging AI/ML to drive impactful decision-making in healthcare.

EDUCATION

Syracuse University, Syracuse, NY — August 2023 – May 2025
Master of Science| Applied Data Science — GPA:3.9/4

Visvesvaraya Technological University, Bangalore, India — August 2017 - August 2021
Bachelor of Engineering| Industrial Engineering and Management — GPA: 8.2/10

EXPERIENCE

Data Engineer, ProHealth Group, Birmingham, Alabama — May 2024 – Present

- Built predictive models (XGBoost) for hospital admissions, discharges, patient risk assessment, improving accuracy by 85%.
- Developed real-time ETL pipelines (AWS Glue, S3, Timestream, Athena) for financial, KPI, and patient diagnosis metrics across multiple branches.
- Implemented causal inference models to assess patient outcomes, optimizing intervention strategies and improving patient engagement by 20%.
- Designed and deployed LLM-powered clinical documentation automation (LangChain, AWS Bedrock), reducing clinician workload by 20%.
- Migrated hospital analytics infrastructure to Snowflake & DBT, enabling real-time insights and reducing query latency by 30%.

Consultant Data Analyst, Flipkart, Bengaluru, India — February 2023 - August 2023

- Reduced product returns by 1.4% through causal analysis of purchase patterns and supply chain adjustments.
- Led A/B testing for EMI adoption using causal difference-in-differences (DiD), driving 16% increase in conversions and 8% incremental revenue/quarter.
- Streamlined KPI dashboards with Google Apps Script, reducing manual effort by 30%.
- Optimized SQL queries and integrated analytics into Flipkart's data platform, improving efficiency and reducing query latency by 50% for faster insights.

Data Analyst, Storeking, Bengaluru, India — October 2021 - December 2022

- Built real-time sales and inventory dashboards (Power BI, Tableau), reducing losses by 2% through optimized demand forecasting.
- Identified high-impact growth regions using counterfactual analysis, increasing sales by 12% MoM.
- Developed supply chain analytics models to improve fulfillment efficiency by 4% and cut procurement costs by 12%.
- Developed vendor performance analytics reports, identifying regional growth opportunities and optimizing supplier management strategies.

SKILLS

Causal Inference & Experimentation: A/B Testing, Difference-in-Differences (DiD), Bayesian Inference, Synthetic Control, Uplift Modeling
Marketing & Analytics: Marketing Mix Modeling (MMM), Media Attribution, ROI Optimization, Customer Segmentation
Machine Learning & AI: TensorFlow, XGBoost, Bayesian Modeling, NLP, Predictive Analytics
Data Engineering & ETL: PySpark, SQL (BigQuery, Snowflake, PostgreSQL), Apache Airflow, dbt, AWS Glue, Redshift
Data Visualization & Business Intelligence: Tableau, Power BI, Looker, Grafana

PROJECTS

Bayesian Marketing Mix Modeling (MMM, Casual Inference, JAX)

- Built a Bayesian Marketing Mix Modeling (MMM) library in Python using NumPyro and JAX, enabling efficient media mix analysis and budget optimization.
- Designed scalable models for media attribution, ROI estimation, and budget allocation, improving marketing decision-making.
- Developed statistical methods for adstock effects, carryover models, and seasonality adjustments, enhancing model accuracy.
- Optimized training and inference time, reducing computational costs by leveraging probabilistic programming techniques.

Cab Dynamic Pricing Model (XGBoost, Scikit-learn, FastAPI, Streamlit, MLflow, Docker, CI/CD Pipelines, AWS Lambda)

- Deployed a dynamic pricing algorithm (85% accuracy) using geospatial and demand data, reducing fare discrepancies by 25%.
- Designed and deployed a Streamlit-based interface, enabling real-time pricing predictions and improving fare transparency for users.
- Optimized retrainable models with user feedback loops, dynamically refining pricing based on demand shifts, historical trends, and surge patterns.
- Integrated real-time monitoring and automated retraining, ensuring continuous model updates and adaptation to evolving ride-hailing market conditions.

RATS – Real-Time Ad Target System (AWS, Kafka, Spark, PostgreSQL)

- Engineered a high-throughput AWS pipeline to process 1TB+ Reddit data using Kafka, Spark, and PostgreSQL for real-time ad targeting.
- Achieved ultra-low latency (600ms) and 3x throughput by fine-tuning Spark configurations and Kafka consumer groups.
- Reduced pipeline latency by 10x, processing 700+ messages/second to deliver real-time ad placements on trending posts.
- Implemented automated scaling and fault tolerance in AWS infrastructure, ensuring high availability and reliability under peak loads.

Cloud-Native Healthcare Data Pipeline (GCP, PySpark, BigQuery, Power BI)

- Built a scalable healthcare data pipeline to process Medicare Part D data efficiently, ensuring high performance and reliability.
- Ingested, cleansed, and transformed raw Medicare Part D data from PostgreSQL into curated, report-ready datasets for analysis.
- Optimized Apache Spark workflows to enhance processing speed while minimizing resource consumption, ensuring cost efficiency and scalability.
- Developed interactive Power BI dashboards for real-time insights into drug utilization and prescriber trends, enabling better data-driven decision-making.