

HARSHA SAI JAGU

(813) 797 9302 | harshasaijagu@gmail.com | [linkedin.com/in/harshasai/](https://www.linkedin.com/in/harshasai/) | github.com/harshasaijagu

PROFESSIONAL SUMMARY

Data Scientist with over 3 years of experience applying predictive modeling, statistical analysis, and machine learning to solve complex business problems in the CPG and BFSI domains. Proven ability to design and analyze large-scale A/B tests, develop and deploy end-to-end ML pipelines, and translate data-driven insights into strategic recommendations for stakeholders. Highly proficient in Python, SQL, PySpark, Big Data and Azure cloud.

WORK EXPERIENCE

Graduate Assistant (Data Scientist)

Jul 2024 – May 2025

University of South Florida

Tampa, Florida

- **Built an automated grading system** for open-ended student answers using text mining and NLP techniques: **TF-IDF**, **sentence embeddings** (e.g., **Sentence-BERT**), and **cosine similarity** to compare answers to reference solutions, achieving high agreement with instructor evaluations. Integrated text preprocessing (**tokenization**, **lemmatization**, **stop word removal**) to improve reliability.
- Instructed **300+ students** in Big Data, Application Development for Analytics developing hands-on modules and exercises using Hadoop, Spark, Docker, HTML, CSS, Chart.JS, .NET and MVC architecture.

Data Scientist

Aug 2021 - Jul 2023

Fractal Analytics

India

- **Partnered with business teams** to design and validate KPIs using historical retail data, identifying top drivers of in-store sales performance.
- **Designed and executed an A/B testing framework** with compliance scoring across test vs. control stores, resulting in **40% uplift in sales** through targeted in-store executions.
- **Architected and automated scalable ETL pipelines** on Azure Databricks (PySpark), processing **100M+ retail records** (TRAX) into a Medallion architecture (Bronze–Silver–Gold) to support KPI modeling and downstream analytics.
- **Scaled the solution globally** by developing modular, object-oriented **Python/PySpark** pipelines and end-to-end Azure Data Pipelines, enabling seamless **onboarding of new countries, retailers, and channels**.
- **Built a Top Drivers Analysis** module using machine learning models (XGBoost) to uncover KPIs most influencing customer engagement.
- **Developed a Customer Journey module** leveraging **funnel analysis** of clickstream data to pinpoint drop-off points and reduce friction in the digital journey.
- **Designed a real-time Alerts module** (anomaly detection, change point detection, trend reversals, seasonality shifts) using **statistical methods** (ruptures, STL decomposition), enabling proactive responses to unusual customer behavior.
- Recognized with the **Star Award - Good Samaritan** (Aug 2022) for outstanding performance and exceptional commitment.

EDUCATION

University of South Florida, Muma College of Business

Aug 2023–May 2025

Master of Science, Artificial Intelligence and Business Analytics, **GPA: 4/4**

Tampa, Florida

Indian Institute of Technology Indore

Aug 2017–May 2021

Bachelor of Technology, Mechanical Engineering

India

TECHNICAL SKILLS

- **Programming & Data Manipulation:** Python, Pandas, NumPy, PySpark, SQL, R, C#, C++, Data Structures & Algorithms
- **Machine Learning & AI:** Predictive Modeling, Deep Learning, Clustering/Segmentation, Time Series Forecasting, Anomaly Detection, Top Driver Analysis, Computer Vision, NLP (BERT Embeddings, TF-IDF), Generative AI (RAG, LLMs)
- **Statistics & Experimentation:** Hypothesis Testing (Parametric & Non-parametric), A/B Testing, Causal Inference, Statistical Modeling
- **Frameworks & Libraries:** Scikit-learn, Keras, XGBoost, Prophet, TensorFlow, MLflow, Spark ML, Hugging Face Transformers, .NET
- **Cloud & Deployment:** Azure (Azure ML, ADLS, ADF), Databricks, Docker, Streamlit, FastAPI, Flask, CI/CD (Azure Repos), Git, JIRA
- **Databases & Visualization Tools:** MySQL, PostgreSQL, Cassandra, Power BI, Tableau, Looker Studio, Alteryx

ACADEMIC PROJECTS

- **RAG-Driven Q&A Chatbot:** Engineered a low-latency RAG system with FAISS, LangChain, and GPT-4, deployed via FastAPI and Docker; integrated into a Streamlit UI for real-time, interactive querying.
- **Car Resale Marketplace:** Built ML models (Stacking Regressor + Bayesian optimization) for price prediction and an item-item similarity engine for recommendations; deployed an end-to-end Streamlit app enabling live price forecasts and personalized suggestions.
- **Hybrid Recommendation Algorithm:** Researched and implemented a hybrid recommendation algorithm combining collaborative filtering with BERT embeddings, improving Mean Average Precision (MAP@3) by 29.3% over baseline models.
- **BudgetPulse Web App:** Developed a financial management platform using C# .NET (MVC) with PostgreSQL on AWS, implementing interactive dashboards for expense tracking and spending forecasts; deployed on Azure for high availability and performance.

CERTIFICATIONS

- **Microsoft Certified:** Azure Data Scientist Associate
- **Fractal Analytics Certified:** Data Scientist & SQL Developer Professional
- **Analytics Vidhya Certified:** Time Series Forecasting using Python