SATHWICK KIRAN M S

Contact: +1 (716) 359-1432 | sathwickkiran04@gmail.com | LinkedIn | GitHub | Portfolio | Buffalo, NY

EDUCATION

University at Buffalo – State University of New York — Buffalo, NY
Master of Science in Data Science — *Aug 2024 – Dec 2025*
B.M.S. College of Engineering — Bengaluru, KA
Bachelor of Technology in Computer Science and Engineering — *Jul 2018 – May 2022*

SKILLS & Certifications

Certifications: Microsoft Azure Data Science Associate (DP-100), AWS Cloud Practitioner, Snowflake Data Warehousing, Databricks Fundamentals, CompTIA, Python, Gitlab, Certification in Cybersecurity from University of Maryland, Linux CLI
Programming: Python, R, SQL, C++, Java, Scala Pyspark
Cloud/MLOps: AWS, Azure, Terraform, Docker, Kubernetes, CI/CD, Hadoop, GCP, Snowflake
Tools: PowerBI, Tableau, GitHub, GitLab, MATLAB, Databricks, Excel
Scripting & Automation: Shell, YAML, Git
Soft Skills: Problem Solving, Communication, Cross-Functional Team Collaboration, Stakeholder Collaboration, Data-Driven Decision Making, Scalability, Actionable Insights
Others: Generative AI, LLM fine-tuning, FastAPI, PyTorch, TensorFlow, NoSQL (MongoDB), Hive, Scala, Elasticsearch, GraphQL, Business Consulting, GTM, Metabase, Azure Synapse Analytics, Data Flows, Datalakes, Parquet/Delta, KQL (Basics), Log Analytics (Exposure), Azure DevOps (Exposure)

WORK EXPERIENCE

Sony Corporation* | *Data Engineer | *Bengaluru, KA* — *Aug 2022 – Apr 2024*

- Designed and deployed ETL pipelines in Python and SQL to extract, transform, load, and integrate data from structured and unstructured data sources, ensuring data quality, lineage, and automation of data extraction from product databases.
- Built and containerized data workflows using Docker and Kubernetes, deployed on AWS, enabling scalable, fault-tolerant analytics for team product lifecycle and usage data through automated data pipelines.
- Applied statistical modeling and machine learning techniques to analyze data and support go-to-market strategies.
- Automated reporting pipelines and monitoring systems, reducing manual reporting effort by 15% and enabling real-time dashboards for product feature usage, engagement, and customer value realization.
- Delivered executive dashboards in Power BI, providing analytics and data access through visualizations to support strategic decision-making for stakeholders.
- Integrated GitLab with Jira for CI/CD workflows, cutting task processing time by 20% and boosting team productivity through automated application build and deployment.
- Collaborated across multiple engagements with cross-functional teams including finance, product, and engineering to deliver analytics that informed business consulting initiatives and go-to-market strategies.
- Leveraged relational databases and SQL query authoring for data access, utilized Big Data frameworks for distributed data processing and product usage analysis.
- Partnered with cross-functional teams to translate analytical findings into actionable insights and improve decision-making efficiency.
- Won 3rd place in Sony's "First Challenge" developing a document automation and analytics system that reduced manual effort by ~30%, demonstrating end-to-end automation and engineering expertise.

PROJECTS

AI Multi-Agent System for Short-Selling Prediction — *Nov 2025*

- Developed an **AI multi-agent framework using LangGraph** (Analyst, Model, Risk agents) to autonomously identify short-selling opportunities in the U.S. equities market.
- Processed **5+ GB of financial data** from Kaggle using Python, Pandas, and HDFS on a **Vultr cloud cluster**, enabling distributed ingestion and feature engineering.
- Built a **Logistic Regression ML pipeline** with MLflow tracking, achieving **0.64 AUC and 0.45 F1**, and performed predictive short-signal modeling and evaluation.
- Integrated a **Qdrant vector database** for semantic retrieval and embedded financial indicators (RSI, MA ratio, volatility) to enhance AI-driven reasoning and decision-making.
- Designed and deployed an **interactive Streamlit dashboard** with HuggingFace integration, visualizing short candidates, performance metrics, and automated stock narratives, while managing end-to-end cloud infrastructure using **Hadoop, Vultr, and LangGraph**.

Advanced Data Analysis and Scalable Big Data Processing with Hadoop and NLP Models — *May 2025*

- Built and deployed a **big data pipeline using Hadoop and Spark** to process 3M+ Amazon book reviews, enabling large-scale sentiment analysis on unstructured text.
- Developed and tuned **machine learning models (Logistic Regression, SVM, Random Forest, Naïve Bayes)** using TF-IDF features, achieving strong F1 and accuracy performance.
- Conducted **feature importance and error analysis** with Random Forest and SHAP to identify key predictors and diagnose misclassifications from linguistic nuances.
- Visualized **genre-wise and temporal sentiment trends** via Power BI dashboards and explored **Generative AI + RAG workflows** including LLM fine-tuning and vector databases.