SARVESH KRISHNAN RAJENDRAN

sarveshk@bu.edu | Sarvesh Krishnan | Sarvesh

Education

Boston University Sep 2023 - Jan 2025
Master's in Applied Data Analytics
- *Relevant Coursework: Big Data Analytics, Machine Learning, Data Mining, Database Design and Implementation*

Vellore Institute of Technology, Chennai Aug 2019 - May 2023
B.Tech. Mechanical Engineering
- *Affiliations and Certifications: IEEE Automation, Machine Learning at Atom Robotics*

Professional Experience

Data Science Consultant - (Contract) **Binghamton, NY**
Bridge Green Upcycle Mar 2025 - Present
- Developing machine learning models to estimate battery State of Health (SoH) and Remaining Useful Life (RUL), enabling classification for recycling and second-life applications.
- Performing feature extraction like Principal Component Analysis(PCA), Non-negative Matrix Factorization (NMF) and Auto Encoders (AE) from voltage-capacity curves and operational data to reduce model complexity and improve prediction accuracy.
- Deploying scalable Battery Health Monitoring Systems (BHMS) on cloud platforms, integrating real-time data from batteries for predictive analytics and remote asset monitoring.

Data Science Fellow **Boston, MA**
Fellowship.AI Jan 2025 - Mar 2025
- Developed an AI-powered regulatory compliance monitoring system to automate rule validation from financial policy documents and business data using knowledge graphs and RAG-based LLMs.
- Designed graph-based representations of 150+ financial rules and integrated vector search with LangChain and LLMs to dynamically generate executable SQL queries for rule validation.
- Fine-tuned LLM prompting strategies and retrieval mechanisms to minimize hallucination and improve SQL generation accuracy, increasing rule-to-query fidelity by 30%.

Data Science Intern **Chennai, India**
Vestas Wind Systems A/S Sep 2022 - Mar 2023
- Developed LSTM models in Python to forecast turbine component failures using historical SCADA data; integrated outputs with Power BI via daily batch pipelines for visualization and maintenance planning.
- Applied supervised machine learning models (Random Forest, XGBoost and Decision Tree's) to classify maintenance priorities from turbine performance logs, enhancing maintenance efficiency by 15%.

Research Experience

Vellore Institute of Technology **Chennai, India**
Summer Research Intern Jun 2021 - Aug 2021
- Developed a lane detection system using **Hough transform**, **OpenCV**, and **Arduino** for speed calculation, achieving **96.3%** accuracy on straight roads and **90.4%** on curves. Embedded a voice assistant for real-time notifications.
- Engineered speed sensor modules and an automatic alerting system to adjust vehicle speed based on lane conditions, enhancing safety and efficiency. Presented this system at **i-PACT'21**.*[LINK]*

Academic Projects

AWS Spot Price Prediction *[GITHUB]* **BU**, Dec 2024
- Developed a scalable PySpark pipeline to process over 27 million AWS spot pricing records, significantly enhancing time-series forecasting capabilities and enabling predictive analytics for cost-efficient cloud resource usage.
- Trained and evaluated baseline models (Linear Regression, Decision Tree, Random Forest) and a distributed feed-forward neural network using Elephas (Keras + Spark) on GCP Dataproc, with the NN achieving RMSE of 0.021.

Healthcare Provider Fraud Detection *[GITHUB]* **BU**, Dec 2024
- Developed an end-to-end fraud detection system analyzing 500,000+ healthcare claims, integrating Apache Airflow for batch ETL workflows and MLflow for model tracking and versioning.
- Trained and optimized multiple models, including Logistic Regression, Decision Trees, Random Forest (73.61% accuracy), and Neural Networks, utilizing stacking ensemble learning to combine tree-based models with deep learning for fraud detection.

Sentiment Analysis *[GITHUB]* **BU**, Mar 2024
- Developed a real-time sentiment analysis system using PySpark for processing 1.6 million tweets and Apache Kafka for tweet ingestion, utilizing AWS services including EMR, S3, and EC2 for scalable and distributed data handling.
- Enhanced sentiment classification accuracy to 83% by fine-tuning DistilBERT on AWS GPU instances, and built a Streamlit-based user interface on AWS EC2 for live tweet analysis and real-time sentiment prediction.

Hogwarts Chatbot *[GITHUB]* **BU**, Jan 2025
- Built a Retrieval-Augmented Generation (RAG) chatbot by integrating Google Gemini API with Pinecone and designed a memory-aware Streamlit interface supporting multi-turn conversations with dynamic prompt optimization via LangChain.

Skills

Languages & Tools	Python, SQL, PostgreSQL, MySQL, Power BI, PySpark, AWS, Google Cloud Platform
Libraries & Framework	TensorFlow, Scikit-learn, LangChain, Apache Kafka, Apache Spark, Hadoop, Elephas
Technologies	Machine Learning, Deep Learning, Natural Language Processing (NLP), Retrieval Augmented Generation (RAG), Knowledge Graphs, Anomaly Detection, Fraud Detection