Jayasurya Sakthivel

Chandler, AZ | jayasuryasakthivel@gmail.com | linkedin.com/in/jayasurya-sakthivel | github.com/jayasurya3012 | (602) 804-9928

EDUCATION

Master of Science in Data Science, Analytics & Engineering *Aug 2024 – May 2026*
Arizona State University (ASU), Tempe, AZ - **GPA:** 4.00 / 4.00

Bachelor of Technology in Computer Science and Engineering with Specialization in AI & ML *Aug 2020 – May 2024*
Vellore Institute of Technology (VIT), Chennai, TN, India - **GPA:** 3.53 / 4.00

EXPERIENCE

AI/ML Engineer – Vizzda *Aug 2026 – Present*

- Created and deployed end-to-end data pipelines for automated real estate document processing, orchestrating **Google Document AI**, **Tesseract OCR**, and Gemini 2.5 Flash in a 3-source consensus extraction system across Maricopa and Miami-Dade County processes.
- To identify beneficial owners at scale, we created entity resolution and skip-trace systems that use fuzzy matching algorithms, Bizapedia, and the Enformion API with adjusted similarity thresholds, cost-control layers, and multi-hop LLC traversals.
- Developed a modular, multi-county pipeline in Python integrating external APIs and **Playwright browser automation**, LLM-powered structured extraction, and rules-based enrichment to produce contact-ready commercial real estate intelligence daily.

Research Assistant – Ira A Fulton of Engineering, ASU *May 2026 – Present*

- Implemented 6 **federated learning** algorithms combining FedAvg/FedGmd aggregators with SGD, ASS, and SASS trainers, running 432 experiments across IID and non-IID MNIST and medical data settings like REFUGE and LAG with corruption levels up to 80% to benchmark **adaptive step-size methods** under noisy federated conditions.
- Identified why FedGmd hurt accuracy by up to 6% at high corruption ratios rather than helping, tracing it to geometric median collapsing toward the corrupt majority when over 50% of clients are corrupted, and confirmed this theoretically using the RFA paper's breakdown point analysis.

Data Analyst Intern – Amrutanjan Health Care Ltd *May 2023 – July 2023*

- Developed automated **ETL** and reporting pipelines using **PostgreSQL, PySpark** on Databricks, and **Python** to handle 1 million+ sales records, removing manual procedures and saving **300+ staff hours** per year, with anomaly detection checks flagging data errors before executive reporting.
- Engineered **14 demand** and seasonality variables using **DBSCAN clustering** and statistical approaches, creating **LSTM** and **Seasonal ARIMA** forecasting models that achieved accuracy of 90%, increasing procurement planning and lowering stock-out risk.
- Delivered **Power BI** variance **analysis dashboards** to uncover cost optimization opportunities, directly influencing procurement strategy and **operational decision-making** for senior leadership.

PROJECTS

LoanLens - Credit Risk Decision Intelligence, Capstone Link

- To address a 4:1 default imbalance, we trained **XGBoost, LightGBM**, and **Random Forest** on 1.1 million LendingClub records using class-weighted handling, with XGBoost obtaining an **AUC-ROC** of 0.703 and 0.67 recall on held-out test data.
- Integrated **SHAP**-based explainability to expose per-applicant feature contributions and linked live macroeconomic information from the FRED API at inference time to ensure forecasts match current economic conditions.
- Created a full-stack solution with a **FastAPI** inference layer and a React dashboard, as well as a **GPT-powered** conversational interface that converts model outputs into plain-language explanations for borrowers.

Retail Sales Analytics Platform Link

- Delivered a scalable **retail analytics solution**, increasing visibility into sales performance across 4 years of data through designing a BigQuery **data warehouse** and **Power BI dashboard**.
- Improved query efficiency and reduced data scan costs by optimizing large-scale datasets using **star schema modelling** with partitioning and clustering techniques in **BigQuery**.
- Generated actionable insights on customer behaviour and product performance, driving **KPI tracking** and **trend analysis**, via **SQL-based analytics** layer and interactive **Power BI visualizations**.

AI-Powered Industrial 3D Digital Twin Platform, Cognite track Link

- Developed an interactive **3D digital twin** of an industrial plant increasing operational visibility by **modelling 50+ assets** including sensors, and valves, and equipment from 2D SVG diagrams and rendering them in real time using **React Three Fiber** and **Three.js.**
- Enabled **real-time monitoring** and **anomaly awareness** across 20+ sensor streams by integrating **time-series data pipelines**, dynamically updating asset states, and triggering alerts for abnormal readings in sensors, valves, and critical equipment.
- Architected an **AI agent–driven analytics layer** improving decision support by integrating **Gemini-powered querying** with **RAG-based retrieval** of historical maintenance and live data, to explore asset-level insights with explainable responses and citations.

TECHNICAL SKILLS

Data & Analytics: SQL (MySQL, PostgresSQL), Python (Pandas, Scikit-learn, Matplotlib), R, Azure (AI Studio), GCP (BigQuery, Vertex AI), Data Pipelines, Databricks, A/B Testing, ETL/ELT Pipeline, Tableau, Power BI (DAX, Data Modeling).

AI & Machine Learning: ML Modelling, LLMs, RAG, Multi-Agent Systems, Agentic AI, PyTorch, TensorFlow, LangChain, LangGraph, Databricks, NLP, Computer Vision, MLOps, Prompt Engineering, Vision Language Models, Time Series Forecasting, Prompt Engineering.

Tools: Excel, Jupyter Notebook, VS Code, Cursor, Claude Code, Streamlit, Git, REST APIs, FastAPI.

Publications: UWE-Net: A Deep Learning Framework for Underwater Image Enhancement, Procedia Computer Science 2025