



# Thathsara Rajapaksha

Homagama, Sri Lanka | [ntratoofficial@gmail.com](mailto:ntratoofficial@gmail.com) | tel: +94706161762

[linkedin.com/in/Thathsara-Rajapaksha](https://linkedin.com/in/Thathsara-Rajapaksha) | [github.com/NTRajapaksha](https://github.com/NTRajapaksha)

Portfolio : [NTRajapaksha](#)

## Profile

Data Science graduate with **First Class Honours** specializing in ML, AI, and MLOps. Experienced in developing production-ready ML models, building data pipelines (Spark, Kafka), and implementing MLOps workflows. Skilled in model optimization, cloud deployment (AWS, GCP, Azure), and BI visualization (Power BI, Tableau). Strong problem-solving abilities with proven collaboration and stakeholder communication skills.

## Skills

**Languages & Core:** Python, SQL, R, Java, HTML, CSS, JavaScript, Excel

**ML/AI & Data Science:** TensorFlow, PyTorch, Scikit-learn, XGBoost, OpenCV, YOLO, Hugging Face, LangChain, LangGraph, Keras, MLFlow, AutoML

**Cloud & Infrastructure:** AWS (S3, SageMaker, Lambda, Redshift), GCP (Vertex AI, BigQuery), Azure (ML, Databricks), Docker, Kubernetes

**Data Visualization & BI:** Power BI, Tableau, Looker, Matplotlib, Seaborn, Plotly

**Big Data & Engineering:** Apache Spark, Kafka, Hadoop, Airflow, PySpark, Snowflake, MS SQL Server, Flask, FastAPI

## Experience

### AI Engineer Intern

Feb 2025 - Aug 2025

Codebell PVT LTD, Dikwella Rd, Beliatta, Sri Lanka

- Developed end-to-end ML pipelines spanning computer vision (CNN transfer learning), NLP (BERT/DistilBERT for text classification, Sentiment Analysis), and traditional ML (XGBoost, hybrid recommenders, contextual bandits) across all production projects achieving nearly 80-90% model accuracy.
- Optimized models for production deployment using quantization (40% size reduction), distillation, caching strategies, and implemented fairness-aware ML with bias mitigation.
- Delivered production-ready models with comprehensive documentation through cross-functional collaboration, demonstrating capabilities in model development, optimization, ethical AI practices, and stakeholder communication.

### Freelance ML Engineer & Data Analyst

2023 - 2025

Remote / Self-Employed

- Delivered end-to-end ML solutions and statistical analysis projects for diverse clients using Python, R, and MATLAB across domains including predictive modeling, data pipeline development, and business analytics.

### Customer Relations Officer

Dec 2021 - Apr 2022

LB Finance PLC, Delkanda, Sri Lanka

- Handled financial reporting, customer account management, and client communications, developing strong stakeholder engagement and problem-solving skills in a fast-paced financial services environment.

## Projects

### Azure Serverless Credit Risk API (MLOps & Cloud)

- Built end-to-end MLOps pipeline on Azure achieving 0.76 AUC in credit default prediction, deploying LightGBM model as serverless REST API with auto-scaling architecture and cost optimization (zero USD at idle)
- Implemented native model serialization to resolve version dependency conflicts between training and serving environments, delivering sub-200ms inference latency with SHAP explainability for real-time credit decisions
- Tools Used: Python, Azure ML, LightGBM, Azure Functions, Azure Blob Storage, SHAP

### Vertex AI Production ML Pipeline (MLOps & Cloud Infrastructure)

- Architected end-to-end TensorFlow deployment on Vertex AI with Explainable AI integration (Shapley values), achieving sub-200ms inference latency and processing 5M+ flight records from BigQuery
- Engineered hybrid Wide & Deep neural network with feature crosses and categorical embeddings, deploying production-ready REST API with auto-scaling serverless architecture for real-time predictions
- Tools Used: TensorFlow, Vertex AI, BigQuery, Explainable AI (Shapley Values), Python, Google Cloud Storage

#### **Hybrid Credit Risk Engine & Financial Explainability (Risk & Operations)**

- Engineered credit scoring system processing 2.2M+ loan records with Polars streaming, reducing memory overhead by 87% (8GB to <1GB) and achieving 71% AUC with 30.5 KS statistic
- Calibrated LightGBM model to FICO-style scores (300-850) for regulatory compliance, integrating Llama 3.2 for natural language explanations of loan decisions
- Tools Used: Python, LightGBM, Polars, Llama 3.2, Streamlit, Scorecardpy

#### **AI Portfolio Intelligence Platform (Financial Analytics)**

- Architected automated Financial Analysis Engine using LangGraph and Google Gemini with ensemble forecasting (Meta Prophet + Linear Regression), generating risk metrics (Sharpe Ratio, VaR) for 10+ assets while reducing data costs by 90%
- Deployed production-ready Streamlit application via Docker for reproducible financial modeling across environments
- Tools Used: Python, Meta Prophet, LangGraph, Google Gemini API, Docker, Streamlit.

### **Achievements**

---

#### **1st Place - DataXplore 2025** 2025

- Won 1st place in a national-level Data Science competition organized by the Statistics Society of the University of Sri Jayewardenepura, competing across multiple domains including Exploratory Data Analysis, Machine Learning, Model Explainability, Time Series Analysis, and Statistical Reasoning.

#### **Research Publication - ICACT International Conference** 2025

- Published research on urban computing for sustainable university development at NSBM's 2nd International Conference on Advanced Computing Technologies (ICACT), exploring the application of IoT and machine learning to transportation systems, land use optimization, and air quality monitoring.

#### **Top 15 Finalist - Nexora 1.0** 2025

- Nextintel secured a position among the top 15 teams in Nexora 1.0 competition by developing a Retrieval-Augmented Generation (RAG) model for a university-related scenario, demonstrating proficiency in advanced NLP and LLM applications.

### **Education**

---

#### **BSc (Honours) in Data Science - UGC Approved**

National School of Business Management (NSBM) - Homagama, Sri Lanka

- **First Class Honours** (GPA: 3.76/4.0)
- **Coursework:** Machine Learning, Deep Learning, Big Data Analytics, Statistical Methods, Data Visualization, Cloud Computing, Natural Language Processing
- **View Transcript**

#### **GCE Advanced Level (Physical Science Stream - Mathematics)**

Mahinda Rajapaksha College - Homagama, Sri Lanka

- Com. Maths : C , Physics : S, Chemistry: S, English : B

### **Additional Courses**

---

**MLOps Specialization** - Duke University | **Microsoft AI & ML Engineering Professional Certificate** - Microsoft | **Advanced Machine Learning on Google Cloud Specialization** - Google | **Google Advanced Data Analytics Professional Certificate** - Google | **Microsoft Power BI Data Analyst Professional Certificate** - Microsoft | **Deep Learning Specialization** - DeepLearning.AI | **Machine Learning Specialization** - Stanford University

\*References available upon request