# Thathsara Rajapaksha

Homagama, Sri Lanka | ntratofficial@gmail.com | tel: +94706161762 | linkedin.com/in/Thathsara-Rajapaksha
github.com/NTRajapaksha | Portfolio : NTRajapaksha

## Profile

Data Science graduate with **First Class Honours**, specializing in data analytics, machine learning, computer vision, NLP, and MLOps. Hands-on experience developing production-ready ML models, building data pipelines with Spark and Kafka, and implementing MLOps workflows. Skilled in model optimization, cloud deployment (AWS, GCP, Azure), delivering end-to-end ML solutions, and creating interactive BI dashboards with Power BI and Tableau. Strong problem-solving abilities with experience in collaborative environments and stakeholder communication.

## Skills

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

**ML/AI Frameworks:** TensorFlow, PyTorch, Scikit-learn, OpenCV, MLFLow, Hugging Face, LangChain, LangGraph, LlamaIndex, PySpark, Keras

**Backend Development:** Flask, FastAPI, RESTful APIs

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

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

**Big Data & Distributed Computing:** Apache Spark, Hadoop, Kafka, Airflow, Snowflake, MS SQL Server

**Development Tools:** Jupyter Notebook, Git, Docker, VS Code, Anaconda, Streamlit, CI/CD Pipelines

## Experience

**AI Engineer Intern** Feb 2025 - Aug 2025
Codebell PVT LTD, Homagama, 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

- Independently delivered end-to-end machine learning solutions for diverse clients across multiple domains including computer vision, recommendation systems, agentic AI applications, and statistical analysis projects using Python, R and MATLAB.
- Designed and trained custom ML models for computer vision tasks (image classification, object detection), developed RESTful APIs using Flask/FastAPI for model deployment, and built agentic AI solutions utilizing LLMs, LangChain, and prompt engineering techniques.
- Managed full project lifecycle from requirement gathering to deployment, providing actionable insights through detailed technical reports and maintaining clear client communication throughout project delivery.

**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.

## 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 10 Finalist - Nexora 1.0** 2025

- Secured a position among the top 10 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

**GCE Advanced Level (Physical Science Stream - Mathematics)**
Mahinda Rajapaksha College - Homagama, Sri Lanka

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

## Projects

**IoT Energy Consumption Anomaly Detection with Predictive Analytics**

- Developed a Lambda Architecture-based system for analyzing IoT energy consumption data, implementing batch processing with HDFS and Apache Spark for historical data analysis and real-time anomaly detection using Kafka and Spark Streaming.
- Implemented Linear Regression models with Spark MLlib for energy consumption forecasting and statistical methods (3-sigma rule) for real-time anomaly detection, with results visualized through a Flask-based web dashboard.
- Tools Used: Apache Hadoop, Apache Spark, Apache Kafka, Spark MLlib, Flask, Python, Docker

**Scalable Employee Directory Application on AWS**

- Built a cloud-based employee directory application with automatic scaling and high availability across multiple data centers to ensure zero downtime and handle varying user loads efficiently.
- Implemented secure architecture with database storage, cloud monitoring, and access controls to protect sensitive employee information while maintaining fast performance and reliability.
- Tools Used: AWS (EC2, Application Load Balancer, Auto Scaling, VPC, DynamoDB, S3, CloudWatch, IAM)

**Credit Card Fraud Detection - End-to-End MLOps Pipeline**

- Engineered a robust MLOps workflow for credit card fraud detection, utilizing MLflow for full lifecycle management including automated experiment tracking, metric comparison, and centralized model registry.
- Implemented a containerized deployment strategy using Docker and MLflow Aliases (e.g., `@production`), effectively decoupling model training from inference logic to ensure reproducible production environments.
- Tools Used: Python, MLflow, Docker, Scikit-learn, Flask

**Autonomous Agentic AI Framework**

- Architected a cyclic AI agent using **LangGraph** and **Google Gemini 2.5**, replacing traditional linear chains with a stateful "Reason-Act-Observe" loop for complex problem solving.
- Engineered a **FastAPI** microservice coupled with a **Streamlit** frontend, enabling real-time web research via

**Tavily AI** to bypass LLM knowledge cutoffs (accessing 2025 data).

- Deployed a fully containerized environment using **Docker** and GitHub Codespaces, ensuring reproducible MLOps workflows.
- **Tech Stack:** Python, LangGraph, Google Gemini API, Tavily Search, FastAPI, Streamlit, Docker.

**Colombo Apartments Pricing Data Analysis**

- Conducted comprehensive exploratory data analysis on Colombo's rental apartment market, analyzing 250+ property listings to identify key price determinants including size, location, and amenities.
- Applied statistical methods to detect outliers, analyze correlation patterns, and develop data visualizations that revealed unexpected pricing trends, such as the counterintuitive relationship between furnished status and rental prices.
- Tools Used: Python, Pandas, NumPy, Matplotlib, Seaborn, Jupyter Notebook

**AI Portfolio Intelligence Platform**

- Architected a neuro-symbolic AI agent using **LangGraph** and **Google Gemini** to orchestrate autonomous financial analysis, reducing API costs by 90% via a hybrid execution mode.
- Implemented an ensemble forecasting engine combining **Meta Prophet** and Linear Regression to predict stock prices with risk metrics (Sharpe Ratio, VaR) for 10+ assets simultaneously.
- Deployed a production-ready **Streamlit** application using **Docker** containers to ensure reproducibility across local and cloud environments.
- **Tools Used:** Python, LangGraph, Meta Prophet, Google Gemini API, Docker, Streamlit

**Collaborative Dashboard Development for Business Decision-Making**

- Translated the business requirements into interactive dashboards using Power BI and Tableau, enabling real-time tracking of key performance indicators (KPIs). Improved decision making efficiency for stakeholders by delivering intuitive and actionable visualizations.
- Tools Used: PowerBI, Tableau

## Additional Courses

**MLOps Specialization** - Duke University
Comprehensive training in MLOps principles, data management, and production ML deployment. Gained hands-on experience with Amazon SageMaker, AWS, Azure ML, MLflow, and Hugging Face for building end-to-end ML solutions and pipelines. Applied exploratory data analysis, AI pair programming with GitHub Copilot, model training and optimization on cloud platforms, and deployed LLMs and containerized models using ONNX format.

**Microsoft AI & ML Engineering Professional Certificate** - Microsoft
Specialized training in AI/ML infrastructure, Azure-based ML pipelines, model deployment, and intelligent agent development. Covered end-to-end ML lifecycle management using Azure services including Azure ML and data pipeline optimization.

**Google Advanced Data Analytics Professional Certificate** - Google

**Google Business Intelligence Professional Certificate** - Google

**Advanced Machine Learning on Google Cloud Specialization** - Google

**Mathematics for Machine Learning and Data Science Specialization** - DeepLearning.AI

**Microsoft Power BI Data Analyst Professional Certificate** - Microsoft

**Deep Learning Specialization** - DeepLearning.AI

**Machine Learning Specialization** - Stanford University

**AWS Fundamentals Specialization** - AWS

**Generative AI with Large Language Models** - DeepLearning.AI

**Machine Learning in Production** - DeepLearning.AI

**TensorFlow - Keras Bootcamp** - OpenCV University

***References available upon request**