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, AI 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, YOLO, MLFLow, Hugging Face, LangChain, LangGraph, 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, 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

- 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 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

Projects

Strategic Customer Intelligence & Segmentation

- Developed a hybrid customer intelligence system combining Unsupervised Learning (K-Means) for customer segmentation and Supervised Learning (XGBoost) for predictive behavioral analysis
- Analyzed customer data to identify high-value segments, enabling the marketing team to launch precision targeting campaigns with a 90% AUC score.
- Translated segmentation results into actionable business strategies, directly supporting customer retention and revenue optimization efforts.
- Tools Used: Python, XGBoost, K-Means, Scikit-Learn, Looker, Pandas.

IoT Energy Data Analytics (Big Data & Cloud)

- Architected a Big Data pipeline to analyze high-volume energy consumption data, implementing a Lambda Architecture for both batch (historical) and real-time analytics.
- Utilized Apache Spark and Kafka to process streaming data, detecting usage anomalies and trends that informed energy-saving strategies.
- Deployed the analytical engine on a cloud-ready infrastructure, demonstrating capability in handling large-scale datasets typical of enterprise environments.
- Tools Used: Apache Spark, Kafka, Hadoop, Python, SQL.

Credit Card Fraud Detection (Risk & Operations)

- Engineered a robust fraud detection workflow to identify anomalous financial transactions, utilizing MLflow for experiment tracking and model governance.
- Focused on risk mitigation by implementing a containerized model scoring service, ensuring real-time fraud alerts were delivered to the fraud prevention team with high accuracy.
- Generated automated performance reports to validate model reliability against strict financial compliance standards.

- Tools Used: Python, MLflow, Scikit-learn, Docker, Pandas.

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).
- **Tech Stack:** Python, LangGraph, Google Gemini API, Tavily Search, FastAPI, Streamlit, Docker.

AI Portfolio Intelligence Platform (Financial Analytics)

- Architected an automated Financial Analysis Engine using LangGraph and Google Gemini to orchestrate market research, reducing data retrieval costs by 90% via a hybrid execution mode.
- Implemented an ensemble forecasting model combining Meta Prophet and Linear Regression to predict stock prices with critical risk metrics (Sharpe Ratio, VaR) for 10+ assets simultaneously.
- Deployed a production-ready Streamlit application using Docker, ensuring reproducible financial modeling across local and cloud environments.
- Tools Used: Python, Meta Prophet, LangGraph, Google Gemini API, Docker, Streamlit.

Colombo Real Estate Market Analysis (Business Insights)

- Conducting comprehensive Exploratory Data Analysis (EDA) on 250+ property listings to uncover key pricing determinants such as location, amenities, and size.
- Identified counterintuitive market trends (e.g., furnished status impact on price) and presented these findings as strategic investment recommendations.
- Tools Used: Python, Pandas, Matplotlib, Seaborn, Jupyter Notebook, Statistical Analysis.

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