Thathsara Rajapaksha

Homagama, Sri Lanka | ntratofficial@gmail.com | tel: +94706161762

linkedin.com/in/Thathsara-Rajapaksha | github.com/NTRajapaksha | Portfolio : NTRajapaksha

Profile

First Class Honours Data Science graduate with a track record of delivering production-ready analytics solutions. I specialize in building ML systems that scale (Azure, GCP), creating dashboards that inform decisions (Power BI, Tableau), and translating complex financial data into clear insights.

Skills

PROGRAMMING & STATISTICAL ANALYSIS: Python, R, SQL, MATLAB, Statistical Modeling, Hypothesis Testing, A/B Testing, Experimental Design

DATA VISUALIZATION & BUSINESS INTELLIGENCE: Power BI, Tableau, Looker, Matplotlib, Seaborn, Plotly, Excel (Advanced), Dashboard Development, Stakeholder Reporting

MACHINE LEARNING & AI: TensorFlow, PyTorch, Scikit-learn, XGBoost, LightGBM, Deep Learning (CNNs, RNNs, Transformers), AutoML, Supervised/Unsupervised Learning, Model Optimization (Quantization, Distillation), Explainable AI (SHAP)

GENERATIVE AI & LLMs: OpenAI GPT, Anthropic Claude, Google Gemini, Meta Llama, AWS Bedrock, RAG Architecture, Vector Databases (ChromaDB, Cosmos DB), LangChain, LangGraph, Prompt Engineering, Agentic AI, Model Fine-tuning (LoRA)

CLOUD & MLOPS: AWS (SageMaker, Lambda, S3), Azure (ML Studio, Databricks, Functions, Cosmos DB), GCP (Vertex AI, BigQuery), Docker, Kubernetes, MLFlow, CI/CD, Model Registry

DATA ENGINEERING: Python, SQL, Apache Spark, Kafka, PySpark, Snowflake, ETL Pipelines, Polars, Big Data Processing

BACKEND & API: FastAPI, Flask, REST APIs, Microservices Architecture

Experience

AI Engineer
Feb 2025 - Feb 2026

Codebell PVT LTD, Dikwella Rd, Beliatta, Sri Lanka

- Developed end-to-end ML pipelines for production projects across computer vision, NLP, and traditional ML domains, achieving 80-90% model accuracy and delivering measurable business value through cross-functional collaboration
- 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, collaborating with product and business teams to ensure models met stakeholder requirements and addressed ethical AI considerations including bias mitigation

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

- Managed customer accounts and financial reporting, developing strong stakeholder engagement in fast-paced financial services environment
- Collaborated with cross-functional teams to resolve customer issues, demonstrating problem-solving abilities under pressure

Projects

Multi-Agent Customer Retention System - Azure (Agentic AI & MLOps)
- Built autonomous AI system preventing customer churn using multi-agent architecture that correlates structured billing data with unstructured complaint logs, calculating dynamic churn risk scores for real-time retention decisions
- Deployed production-ready FastAPI microservice on Azure with Vector RAG, achieving sub-second inference for automated offer negotiation based on customer value and sentiment analysis
- Tools Used: Python, LangGraph, Google Gemini Flash, Azure Cosmos DB, Azure Blob Storage, FastAPI, Docker

E-Commerce Customer Intelligence Platform - Azure (MLOps & Business Intelligence)
- Architected serverless Medallion Architecture (Bronze → Silver → Gold) on Azure Databricks processing 100k+ e-commerce orders, achieving cost optimization through PySpark optimization and Delta Lake ACID transactions

- Delivered demand forecasting model (R² 0.36) and ALS-based recommendation engine after strategic pivot from churn prediction, deploying MLflow-tracked models with Docker and creating Power BI "Whale Detector" dashboard identifying high-value at-risk customers
- Tools Used: Azure Databricks, PySpark, Delta Lake, Spark MLlib, XGBoost, MLflow, Power BI, Docker, Azure Data Factory

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

Credit Card Fraud Detection - MLOps Pipeline (Experiment Tracking & Monitoring)

- Engineered end-to-end fraud detection system with automated experiment tracking via MLflow, model registry management, and production-grade drift monitoring for model performance degradation
- Implemented decoupled microservices architecture with CI/CD pipelines, achieving scalable deployment with SHAP explainability for transparent fraud detection decisions
- Tools Used: Python, MLflow, Docker, Flask, Scikit-learn, SHAP, GitHub Actions

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   Apr 2022 - Nov 2025

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)   2020

Mahinda Rajapaksha College - Homagama, Sri Lanka

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

Additional Courses

Microsoft Azure Data Scientist Associate DP-100 Certification (*In Progress*) |**MLOps Specialization** - Duke University | **Advanced Machine Learning on Google Cloud Specialization** - Google | **Google Advanced Data Analytics Professional Certificate** - Google | **Google Business Intelligence 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