

# (Thomas) Hein Thura

## Machine learning Engineer and Data Scientist

✉ - [thomas.h.thura@gmail.com](mailto:thomas.h.thura@gmail.com)

☎ - (+95)9448000829

📍 - Yangon, Myanmar.

🌐 - [Thomas Hein Thura.com](https://ThomasHeinThura.com)(personal website)

🚀 - [@LinkedIn](#) | [@Kaggle](#) | [@GitHub](#) | [@Medium](#)

### Summary

Google-certified **TensorFlow developer**, machine learning enthusiast and **Linux enthusiast** with a strong background in the medical field, and **1.5 years** of working experience as a Freelance Machine Learning and Data Scientist

### EXPERIENCE PROJECTS

#### Freelance Data scientist and Machine Learning Engineer

November 2021- Present

Location - Myanmar

Successfully solved industry-complex **data storage**, **data modeling**, and **machine learning** problems using machine learning and deep learning techniques.

- Developed an 80% accuracy **fraud detection classification model for local bank sectors**, **optimizing accuracy**, reducing false positives, and addressing data engineering workload.
- Implemented an **automatic ETL pipeline** that lead to reducing the data engineering workload by **10%** and ensure monitoring of real-time data drift, data qualities and model performance.
- Built a **Text Summarization website** as a personal hobby project, leveraging natural language processing techniques and testing with state-of-the-art models.
- Utilized oversampling and undersampling techniques to address **imbalanced datasets**, searching for noise patterns, improving model generalisation by 30%, and achieving 99% accuracy.
- Developed models for **weather temperature detection** and energy price detection using time series analysis.



### Achievement

Google **TensorFlow Developer Certificate** ([@link](#))

### Skills

- |                |                 |           |
|----------------|-----------------|-----------|
| • Scikit Learn | • TensorFlow    | • Pandas  |
| • NumPy        | • Prefect orion | • MLFlow  |
| • SQLite       | • PostgreSQL    | • Docker  |
| • Matplotlib   | • EvidentlyAI   | • Seaborn |
| • GitHub/ Git  | • AWS           | • Python  |
| • C++          | • Linux         | • VIm     |

### Languages

- **English** - Business Level proficiency
- **Myanmar** - Native

### Education

**Final year medical student** studied at the University of Medicine 2, Yangon, Myanmar.

## PROJECTS

### ETL pipeline projects

March 2023 - present

- Deploying Machine learning models on the website using **Streamlit** to handle frontend, **MLflow** models data tracking, **EvidentlyAI** to find data drift and using **SQLite** and **PostgreSQL** to store dataset and models, and **Prefect orion** to orchestrate the whole pipeline.

### Singular Website Projects

May 2023 - present

- Deploying a personal website using **streamlit** on **AWS EC2**.

### Kaggle Competition projects and articles

March 2023 - June 2023

- Summit 12 deep learning models and 3 machine learning models in competition, resulting rank top 70%.
- Found out data noise and advanced deep learning models to find the solution.
- Write an article for beginner novice kaggler for sharing experience and knowledge

### Text Summarization website NLP LLM Projects

March 2023- May 2023

- Build **Text Summarization website** as personal hobby projects.
- Testing **Personal AI** website with alpaca Lora 7B model and meta opt 13B model from hugging face

### Proof of Concept projects

January 2023 - March 2023

- **Oversampling** and **Undersampling** on imbalanced datasets (insurance premium pay dataset, credit card fraud dataset).
- Oversampling and undersampling solve the imbalanced dataset problem, which is the poor performance of classifiers on minor classes. And improve 30% of model generalization on the dataset.
- Got 99% accuracy and no FN(false negative) in oversampling with SMOTE testing on Stacking, and ANN models.'

### TensorFlow Projects

November 2021 - January 2023

- **CNN** projects - cifar10 classification, sign language classification, GI tract X-ray segmentation.
- **NLP** projects - IMDb film rating classifications, Reuters documents with articles classification, sarcasm tweets detection, disaster tweet classification, BBC articles classification,
- **Time series** projects - Weather Temperature detection, Energy price detection.

## Soft skills

- communication
- organisation skills
- problem-solving
- time management
- Strong leadership
- task prioritisation skills.
- Generalize real-time problems

## CURRENT PROJECTS

- Personal website
- ETL website
- Projects on UW-Madison GI Tract Image Segmentation,
- Brain tumour detection