Presentation by **Iman Aidi Elham**

PRESERVING CULTURAL HERITAGE SITES THROUGH RANDOM FOREST AND XGBOOST ALGORITHM FOR MICROCLIMATE MONITORING AND PREDICTION

A20EC5006 ✦ Universiti Teknologi Malaysia ✦ 2023

RESEARCH OUTLINE

- Introduction
- Scopes
- Literature Review
- Problem Background
- Methodology
- Objectives
- Research Workflow

INTRODUCTION

Cultural heritage sites are the basis for our global and historical values. They connect us to the traditions left by our ancestors and contribute significantly to the cultural identity of human society (Lombardo et al., 2020).

The preservation of cultural heritage, whether it be buildings or artifacts, is subject to various risks of damage and deterioration that result from microclimate conditions in the surrounding environment. These conditions are determined by several factors, including microclimate parameters such as temperature, humidity, airborne pollutants concentrations, air speed, and others (Fabbri & Bonora, 2021).

PROBLEM BACKGROUND

- Johor Bahru is experiencing frequent climate fluctuations.
- These fluctuations are negatively impacting the aesthetic appeal of heritage sites.
- The tourism industry and local economy are being significantly affected.
- Microclimate changes are causing damage to cultural heritage sites and monuments.
- Balancing consumption and conservation strategies is becoming challenging.
- Preserving cultural heritage and promoting sustainable tourism is a priority.
- This supports cultural heritage tourism and the well-being of communities.

OBJECTIVES

Objective 1

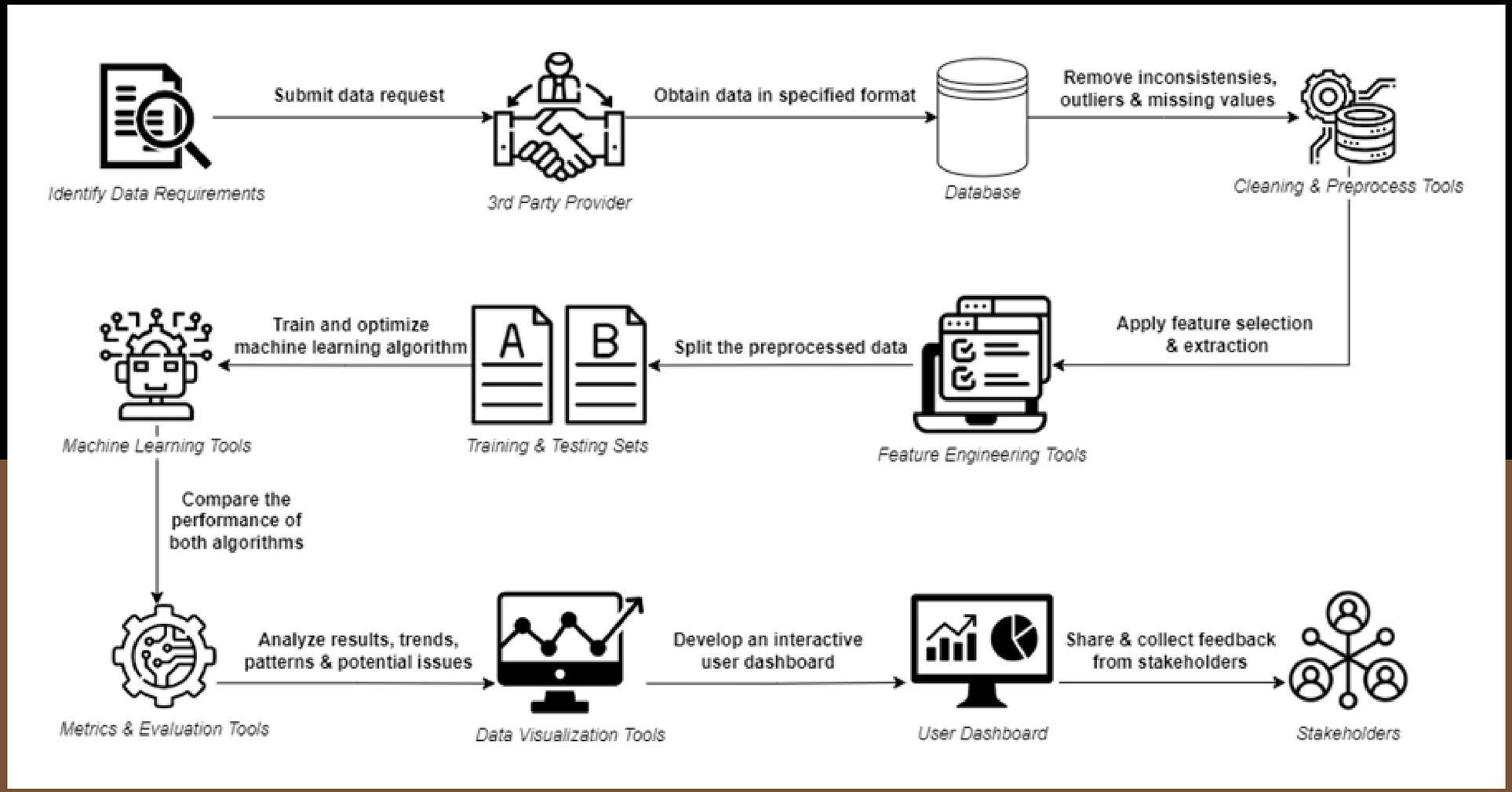
To investigate and identify the most suitable machine learning algorithms for analyzing microclimate data, and to recognize patterns, trends, and potential issues related to the heritage site's preservation.

Objective 2

To evaluate the effectiveness of the developed machine learning models and the dashboard in assisting local authorities to plan preventive maintenance actions that preserve the site's aesthetics and cultural values.

Objective 3

To develop and design a user-friendly dashboard that displays real-time microclimate data and provides recommendations for maintenance actions to assist local authorities with heritage site preservation.



RESEARCH WORKFLOW

Scopes

Scope 1

Research involves obtaining microclimate data from the Malaysian Meteorological Department (MET Malaysia) for a designated heritage site in Johor Bahru. This data contains parameters like temperature, humidity, and wind speed.

Scope 2

The research intends to compare the performance of two different machine learning algorithms between Random Forest and XGBoost to determine the most suitable method for microclimate monitoring and prediction.

Scope 3

The project includes designing and developing an interactive user-friendly dashboard by using data visualization tools that display real-time microclimate data.

Scope 4

The research will involve testing the effectiveness of the developed algorithm and dashboard in assisting local authorities with planning more effective maintenance plans for the heritage site.

RESEARCH AREA ZONE MAPPING

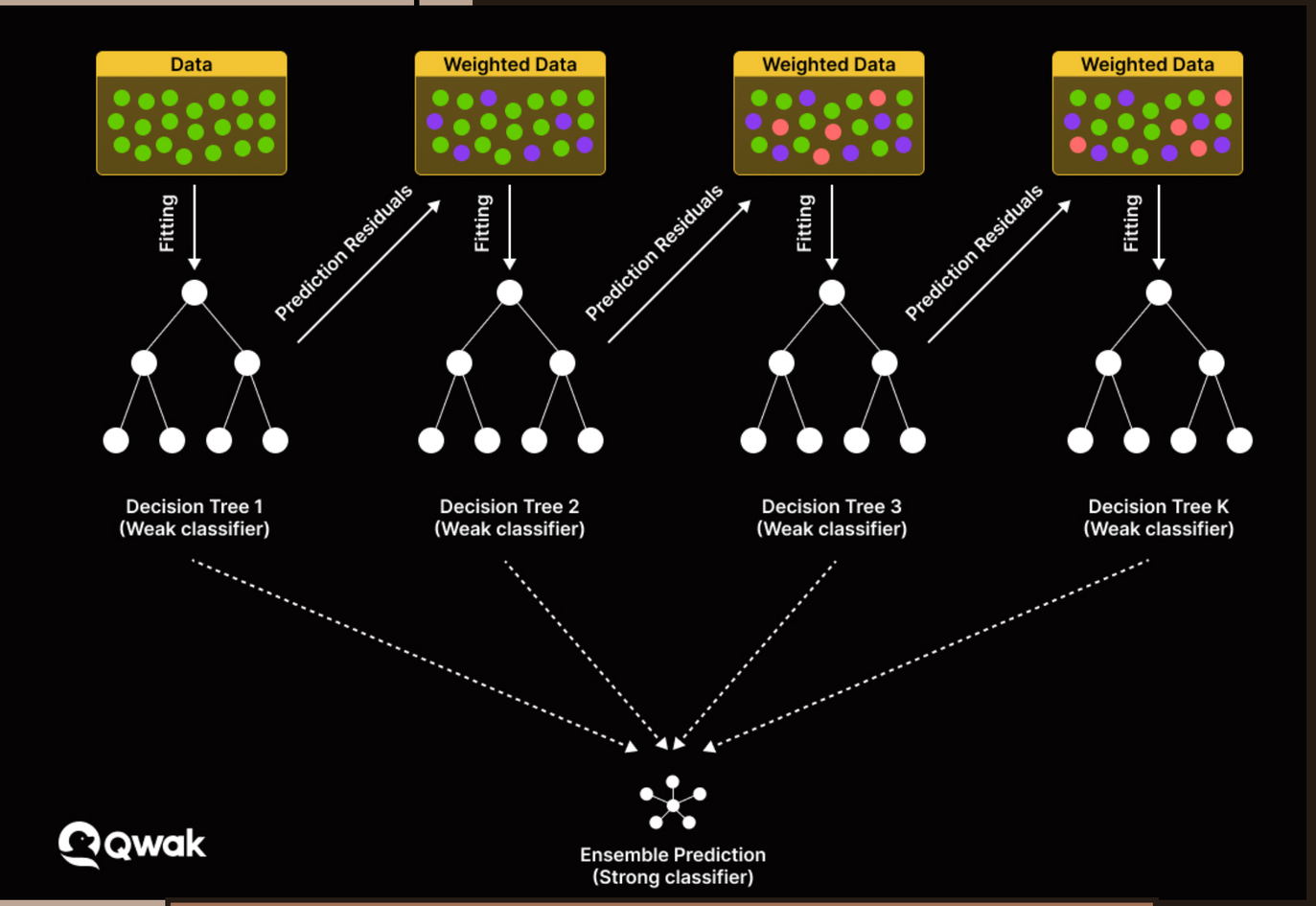
METHODOLOGY

Phase 1	LITERATURE REVIEW
Phase 2	DATA REQUIREMENT & DATA COLLECTION
Phase 3	MACHINE LEARNING MODEL DEVELOPMENT
Phase 4	DASHBOARD DEVELOPMENT

PHASE 1

LITERATURE REVIEW

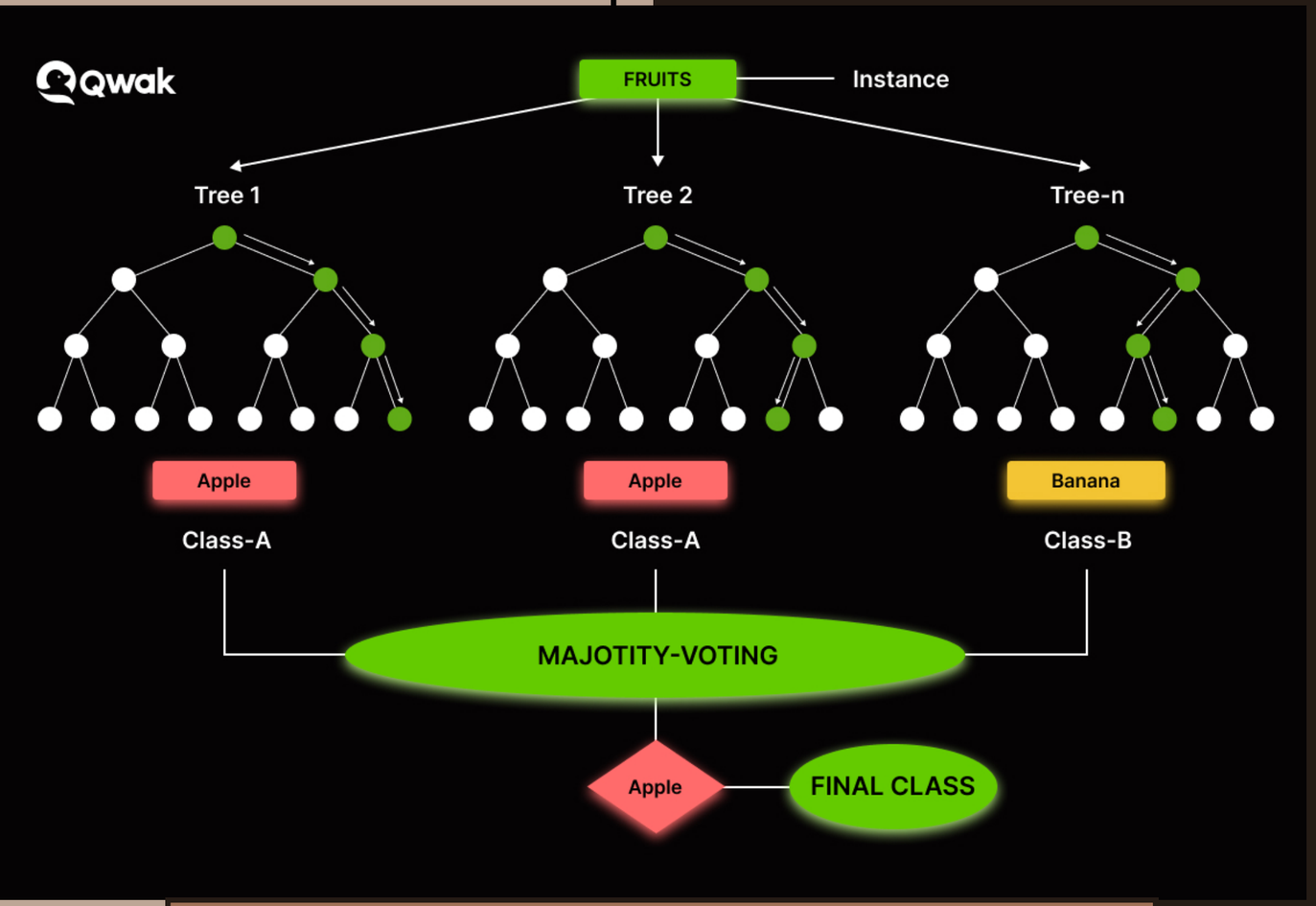
XGBOOST



XGBoost Model Architecture

- XGBoost is a popular machine learning algorithm used for classification tasks.
- Combines multiple weak learners to create a strong model through boosting.
- Handles missing data effectively using surrogate splits.
- Optimized for parallel computing, making it fast and efficient.
- Outperforms other algorithms in terms of accuracy and efficiency.
- Scalable and ideal for handling large volumes of data.
- Enables real-time monitoring and provides insights into environmental conditions.
- Optimized for feature selection to identify influential variables.

RANDOM FOREST



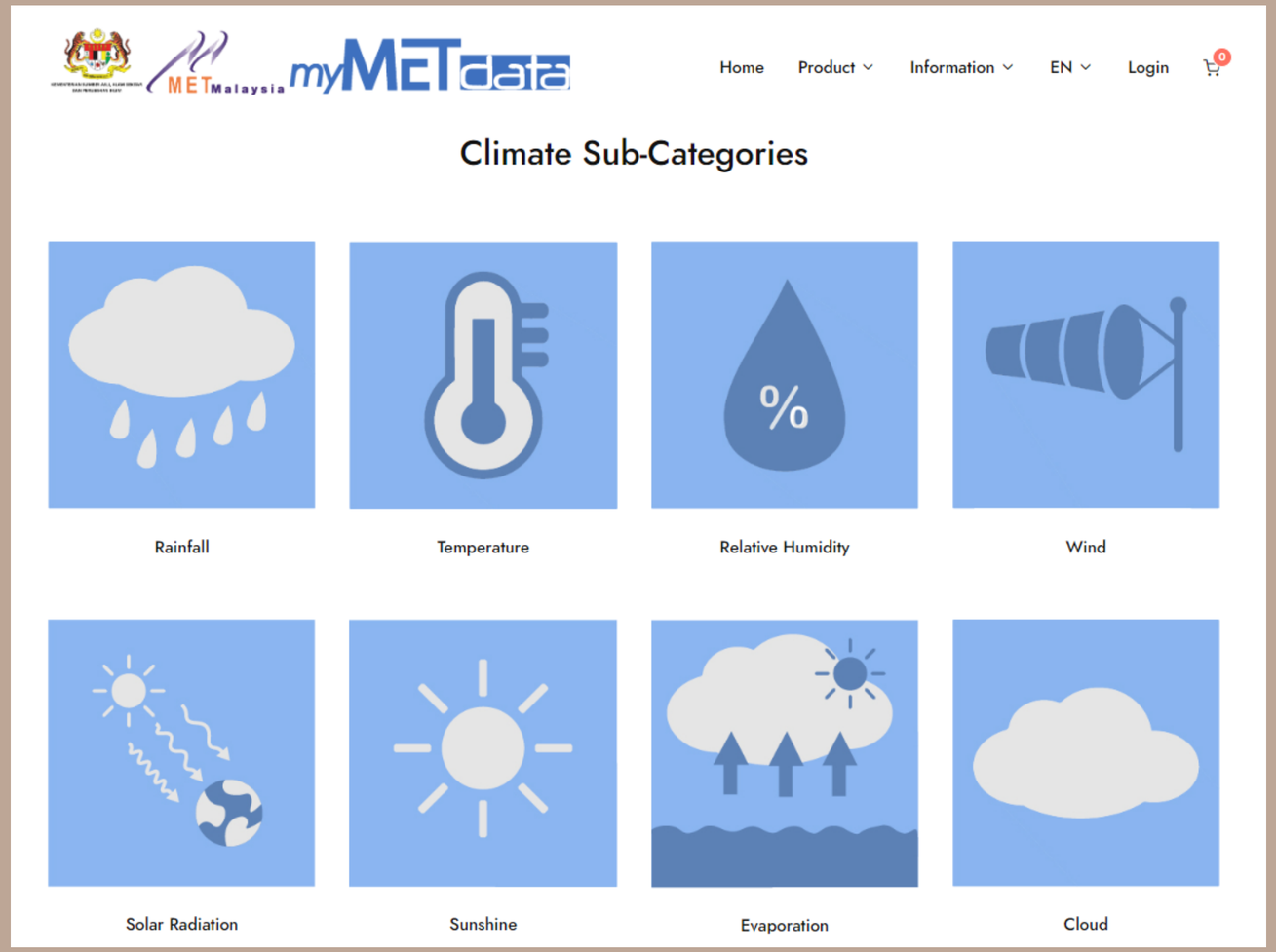
Random Forest Model Architecture

- Random Forest is a popular and versatile ensemble learning algorithm introduced in 2001.
- Can be used for tasks like classification, regression, clustering, interaction detection, and variable selection.
- Uses decision trees and aggregates their predictions to make final classifications.
- Good at handling complex datasets with high dimensionality, noise, and missing data.
- Prevents overfitting and provides robust and high-performance results.
- The algorithm can handle various types of input variables and deal with missing data.
- It can determine the importance of input variables for predictions.
- Random Forest can identify and model interactions between variables, which is useful for microclimate monitoring at cultural heritage sites.

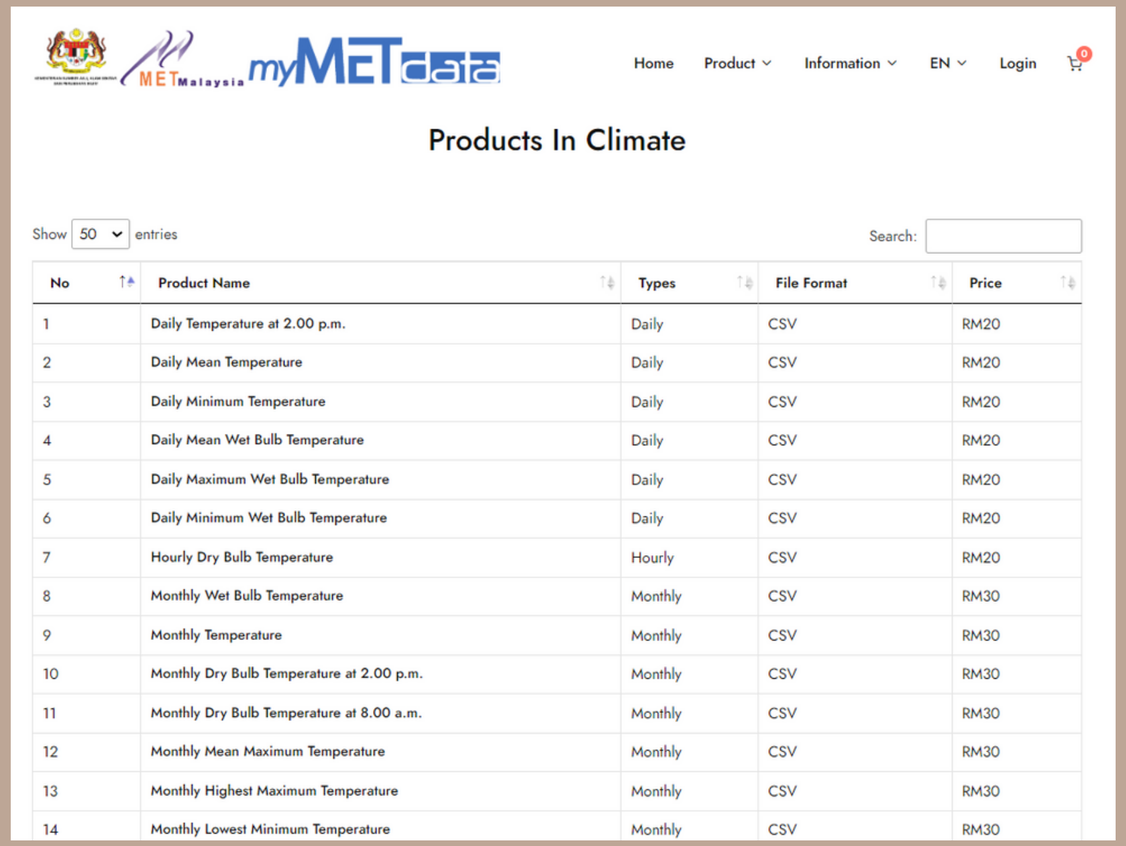
PHASE 2

DATA REQUIREMENT & DATA COLLECTION

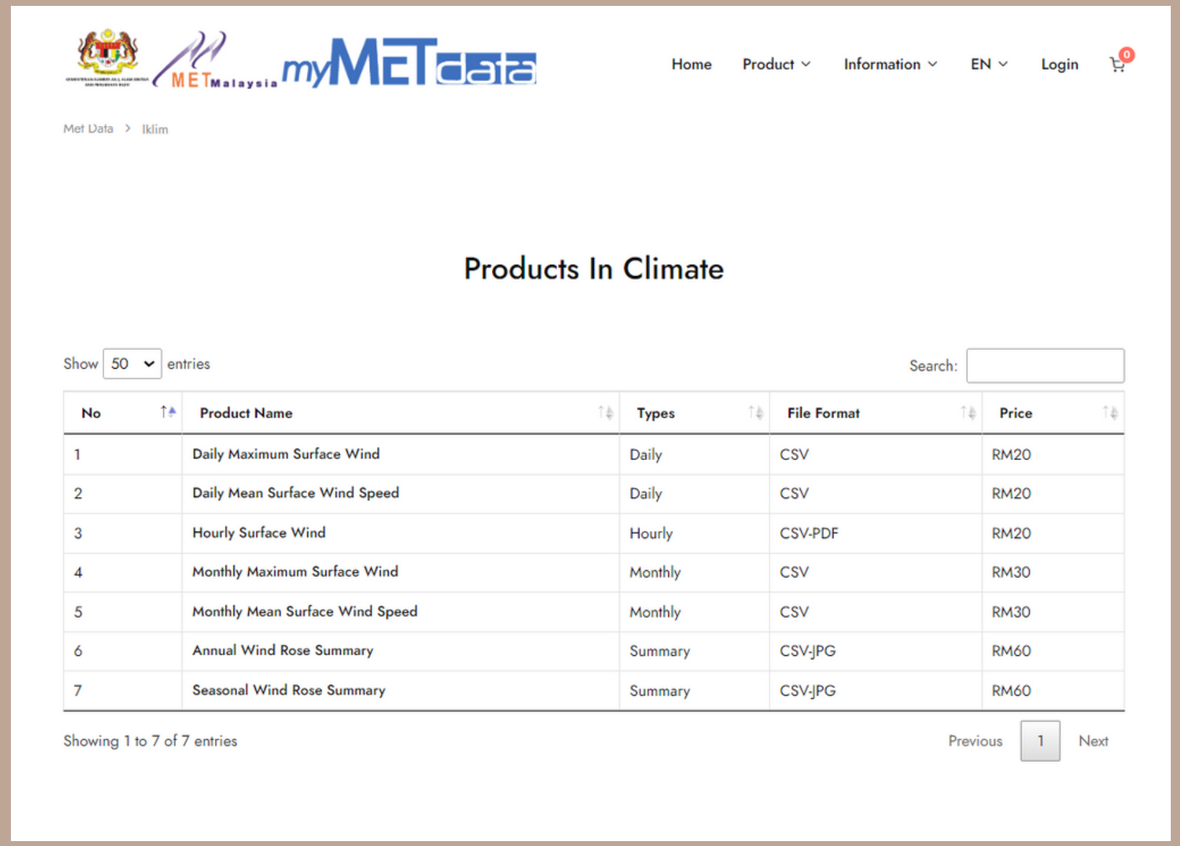
MICROCLIMATE DATA



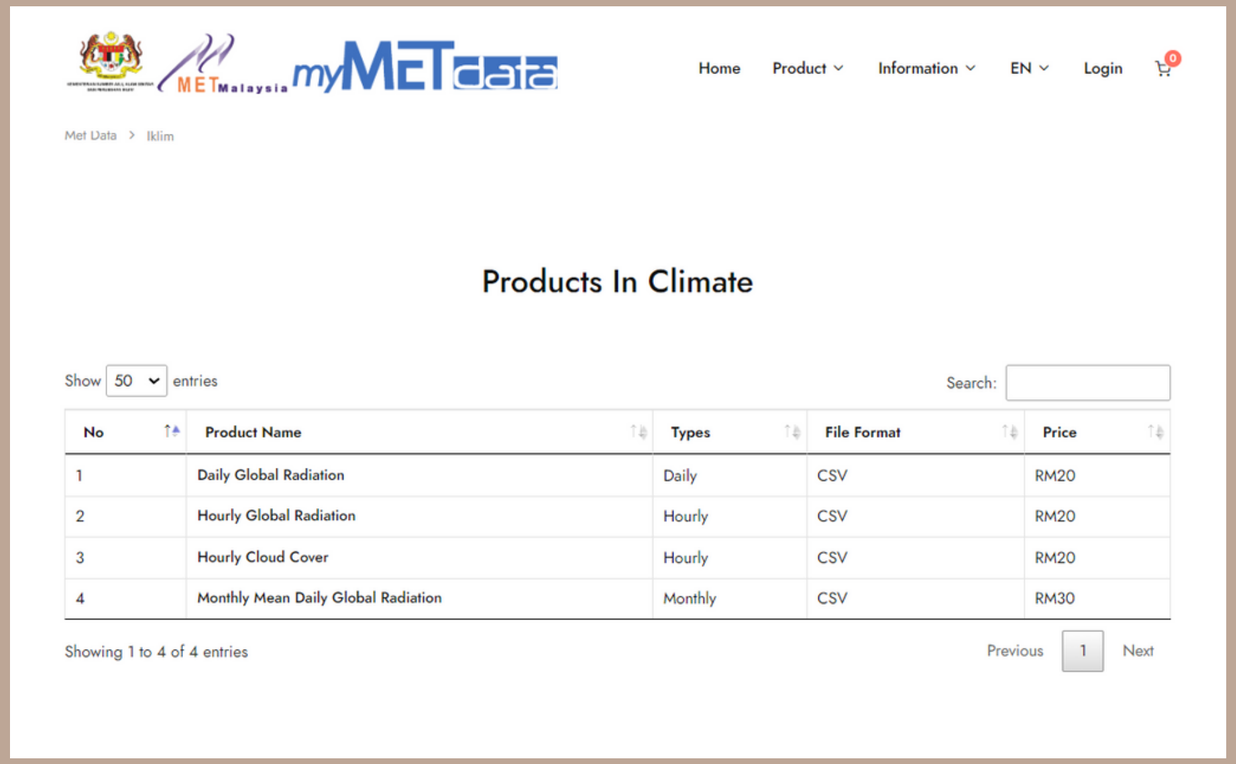
SOURCE : MET MALAYSIA



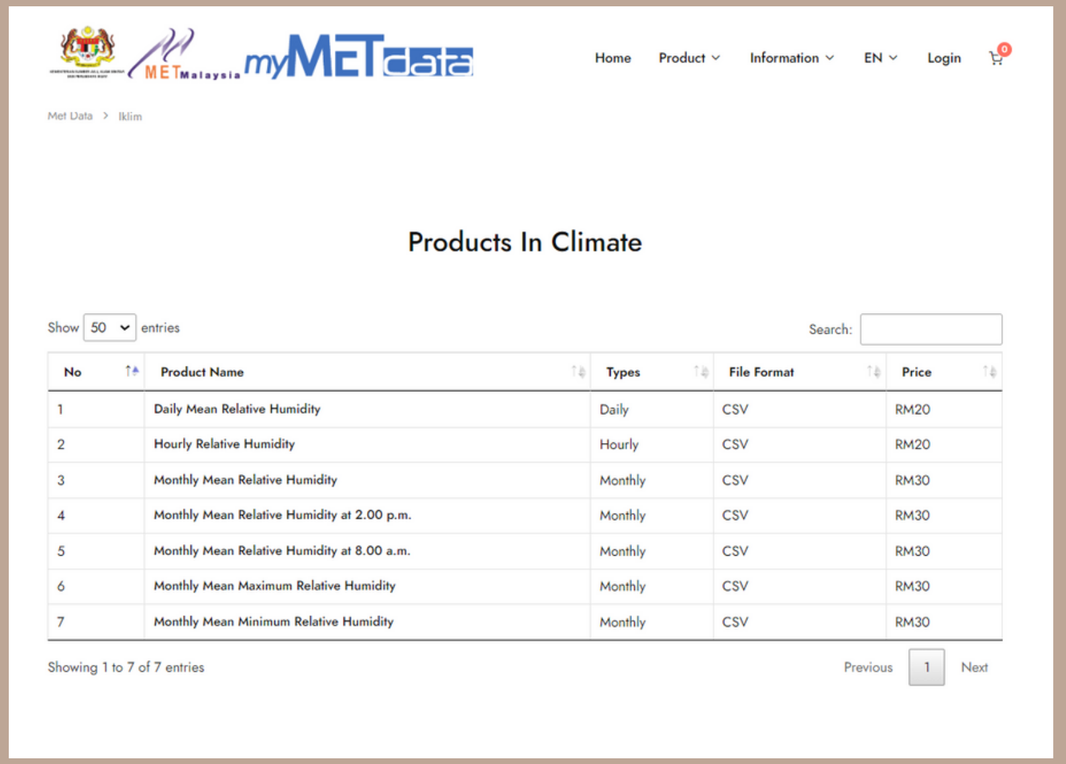
TEMPERATURE DATA



WIND DATA



SOLAR RADIATION DATA



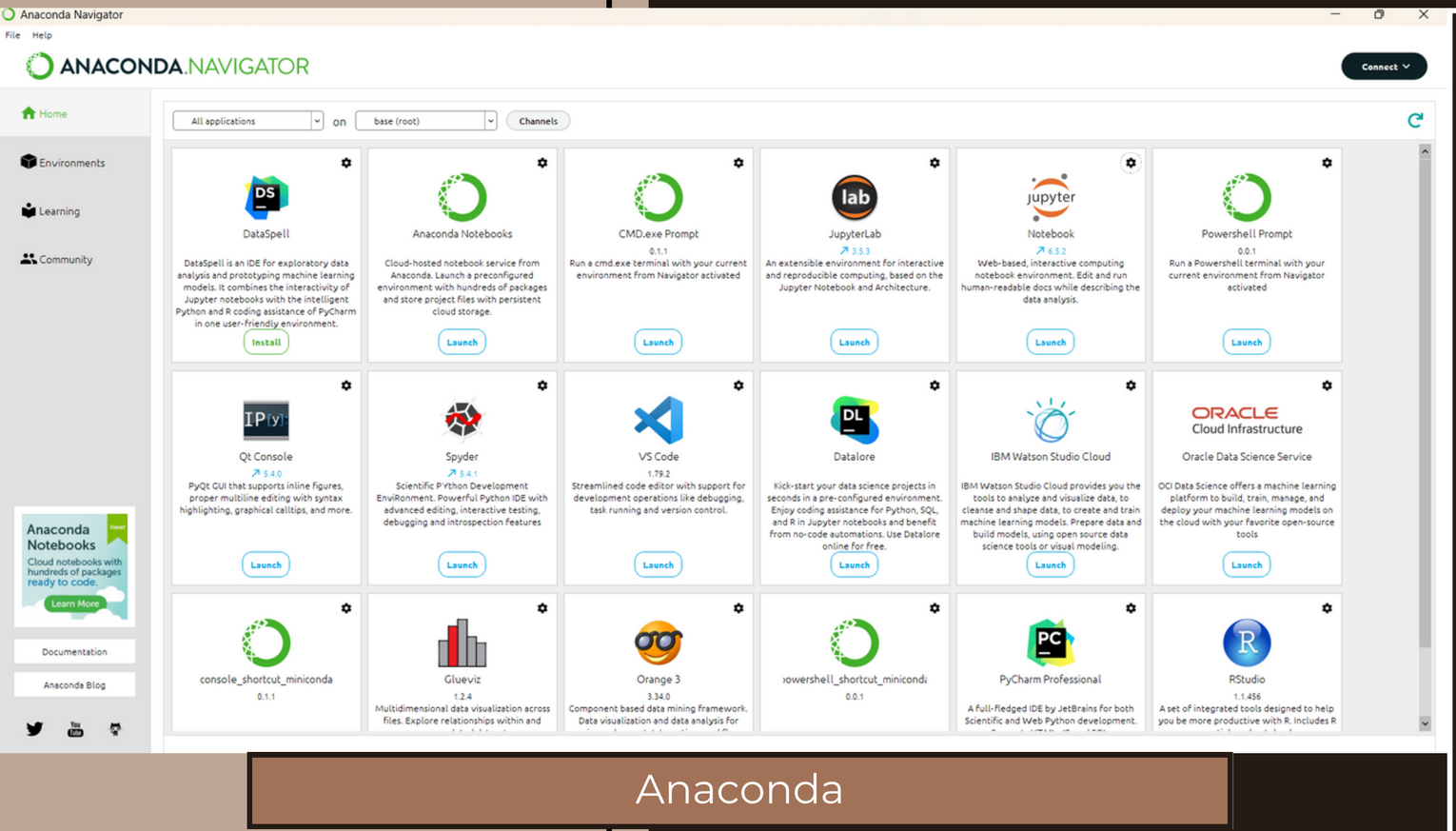
RELATIVE HUMIDITY DATA

PHASE 2

MACHINE LEARNING

MODEL DEVELOPMENT

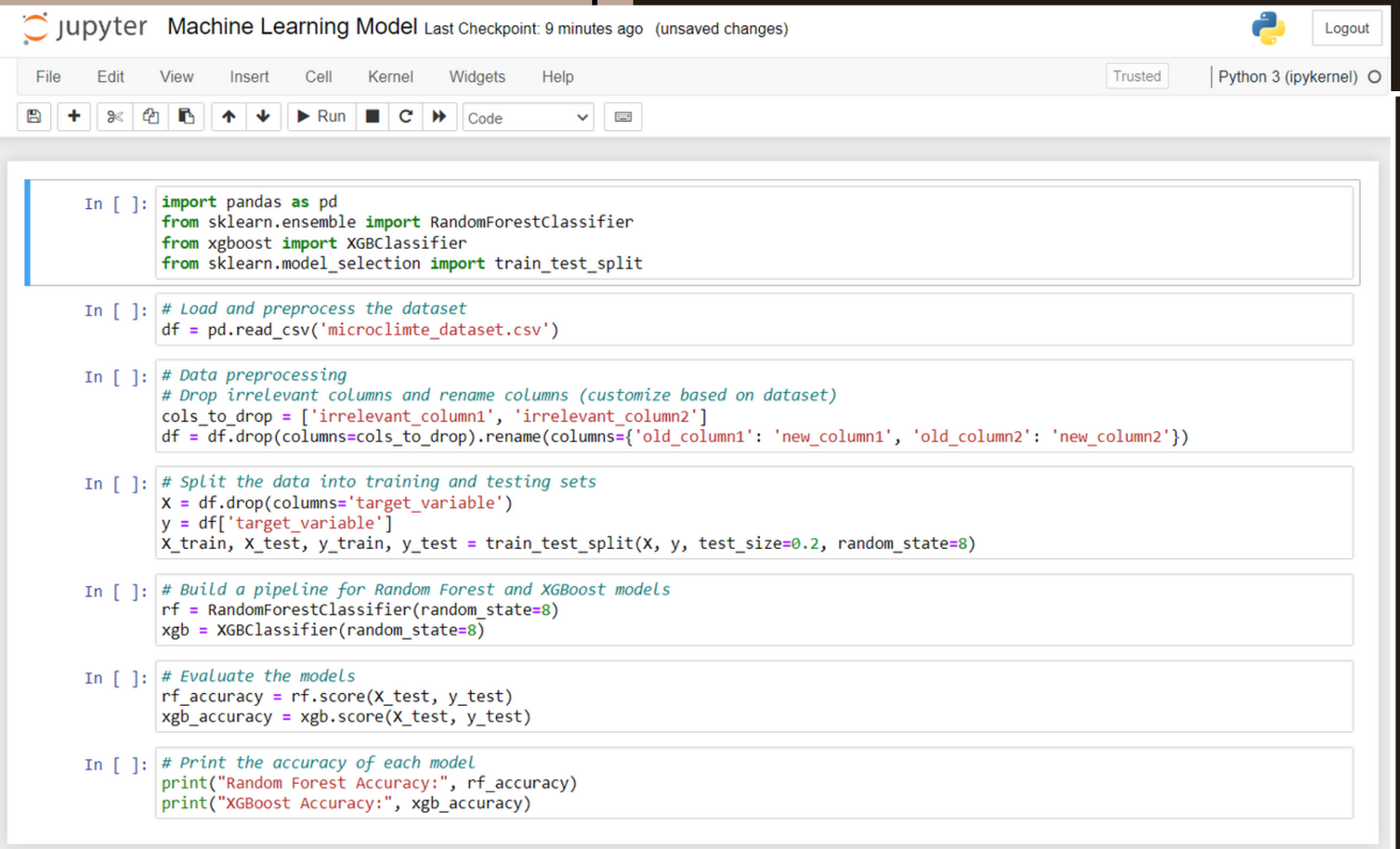
ANACONDA



Anaconda

- Anaconda is a popular distribution of the Python programming language.
- Designed for data science and scientific computing, providing a comprehensive ecosystem of tools and libraries.
- Comes with its own package manager called conda, which simplifies the installation and management of packages and dependencies.
- It includes a wide range of pre-installed packages commonly used in data analysis, machine learning, and visualization.
- Provides an easy-to-use integrated development environment (IDE) called Anaconda Navigator, which allows users to manage environments, install packages, and launch applications.
- Anaconda enables users to create isolated environments, which helps manage different projects with specific package versions and dependencies.

JUPYTER NOTEBOOK



Random Forest Model Architecture

- Open-source web application that allows users to create and share documents containing live code, visualizations, and explanatory text.
- Provides an interactive computing environment, enabling users to write and execute code in small cells or blocks.
- Allows for the combination of code, text, and multimedia elements (such as images and videos) in a single document.
- Users can run individual code cells independently, making it easy to experiment, test, and iterate code in a step-by-step manner.
- Supports the creation of interactive visualizations and data exploration using libraries such as Matplotlib
- Provides a rich set of features, including syntax highlighting, tab completion, and inline output visualization.

PHASE 4

DASHBOARD DEVELOPMENT

DASHBOARD PROTOTYPE

Microclimate Monitoring and Prediction for Preserving Cultural Heritage Sites

Filters

date

Weather Alert

Recommendations

Adjust the HVAC systems

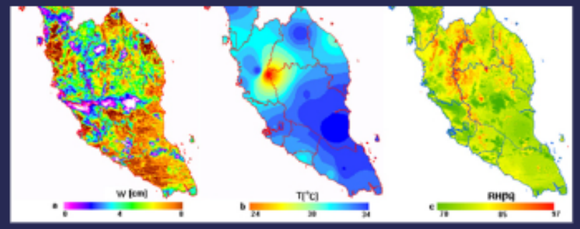
Clock

22:36

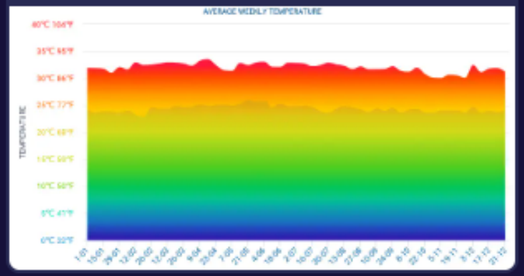
Tuesday
20 Jun 2023

Sites: Johor Bahru High Court

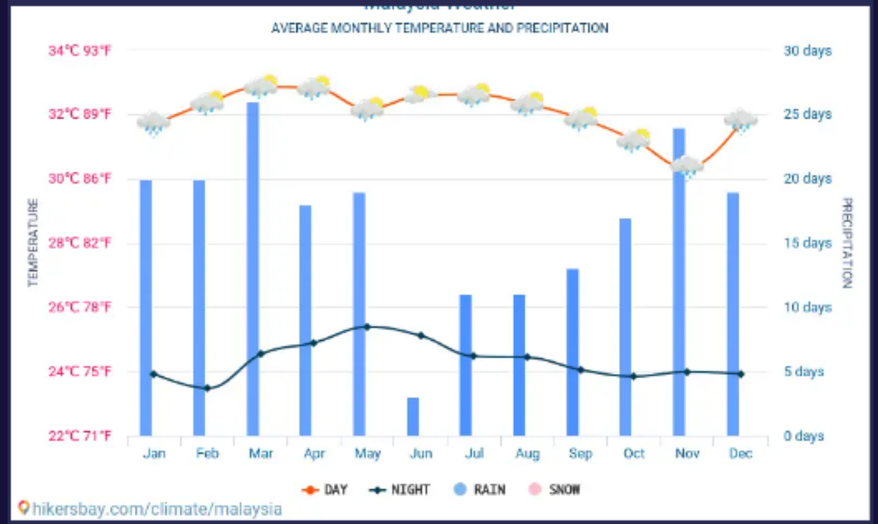
Microclimate Data



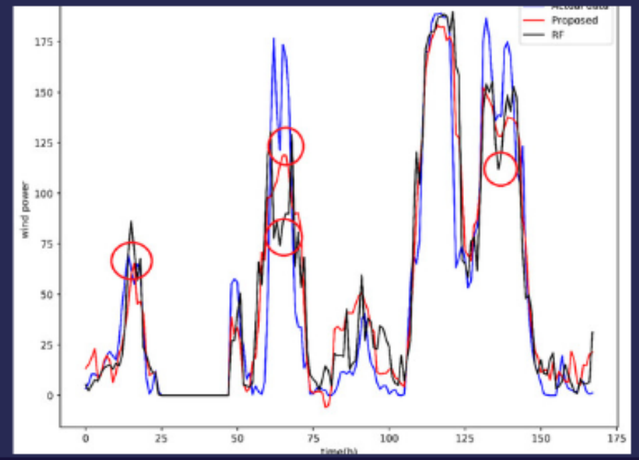
Real-Time Monitoring



Historical Weather Data Analysis



ML Model Accuracy Comparison



Source: MET Malaysia Data

Iman Aidi Elham | PSM1

THANK YOU!