

Presentation by **Iman Aidi Elham**

PRESERVING CULTURAL HERITAGE SITES THROUGH

RANDOM FOREST AND XGBOOST ALGORITHM FOR MICROCLIMATE MONITORING AND PREDICTION

RESEARCH OUTLINE

Introduction

Problem Background

Objectives

Scopes

Methodology

Research Workflow

Literature Review

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.

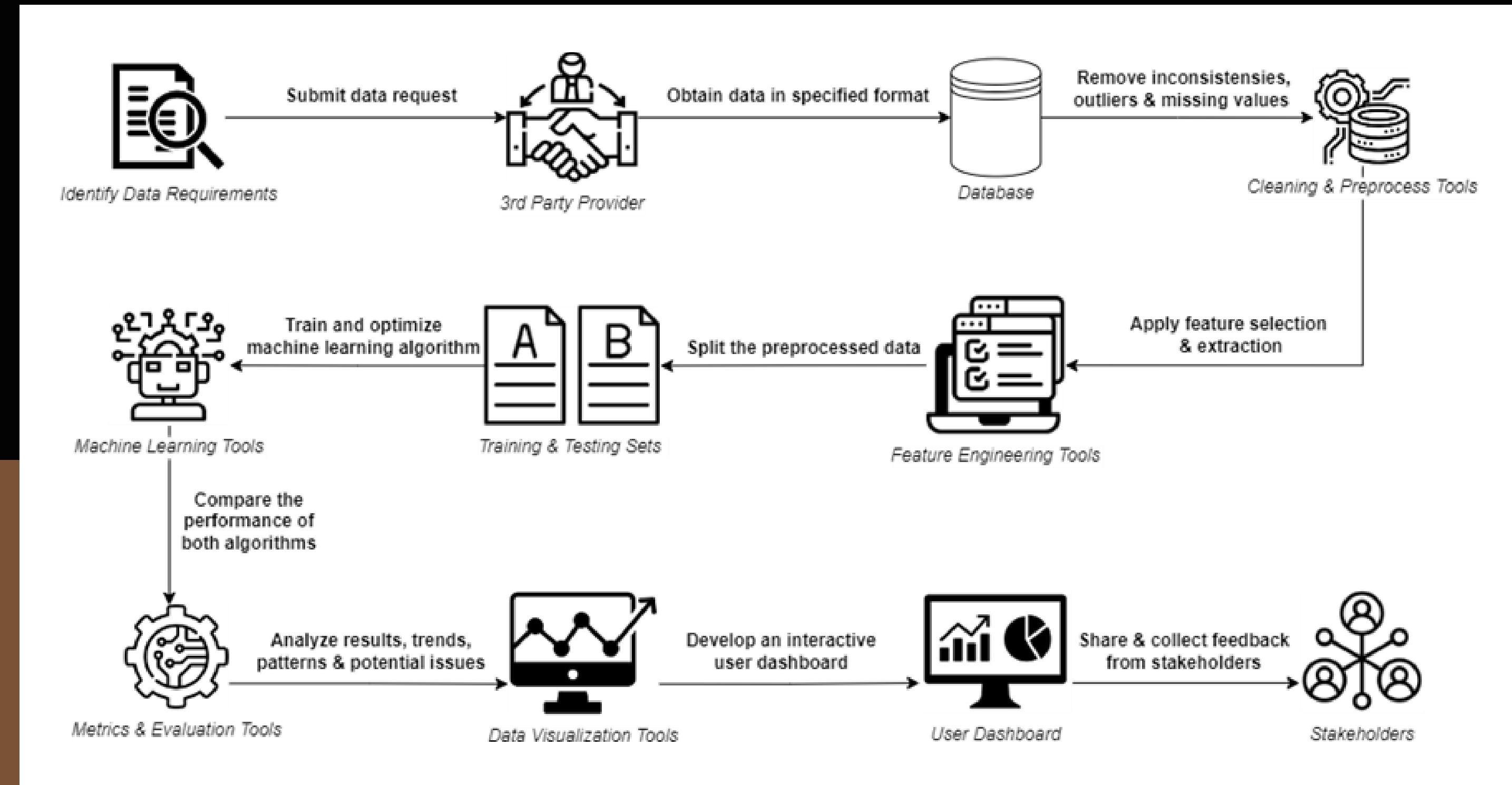
Objective 2

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

— X

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

— X

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

— X

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

Scope 4

— X

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



- Heritage Sites:
1. Sultan Ibrahim Building
 2. Johor Bahru High Court
 3. Sultan Abu Bakar Mosque
 4. Malayan Railway Museum



- Auxiliary Station:
Sultanah Aminah Hospital

METHODOLOGY

Phase 1

LITERATURE REVIEW

Phase 2

**DATA REQUIREMENT & DATA
COLLECTION**

Phase 3

**MACHINE LEARNING MODEL
DEVELOPMENT**

Phase 4

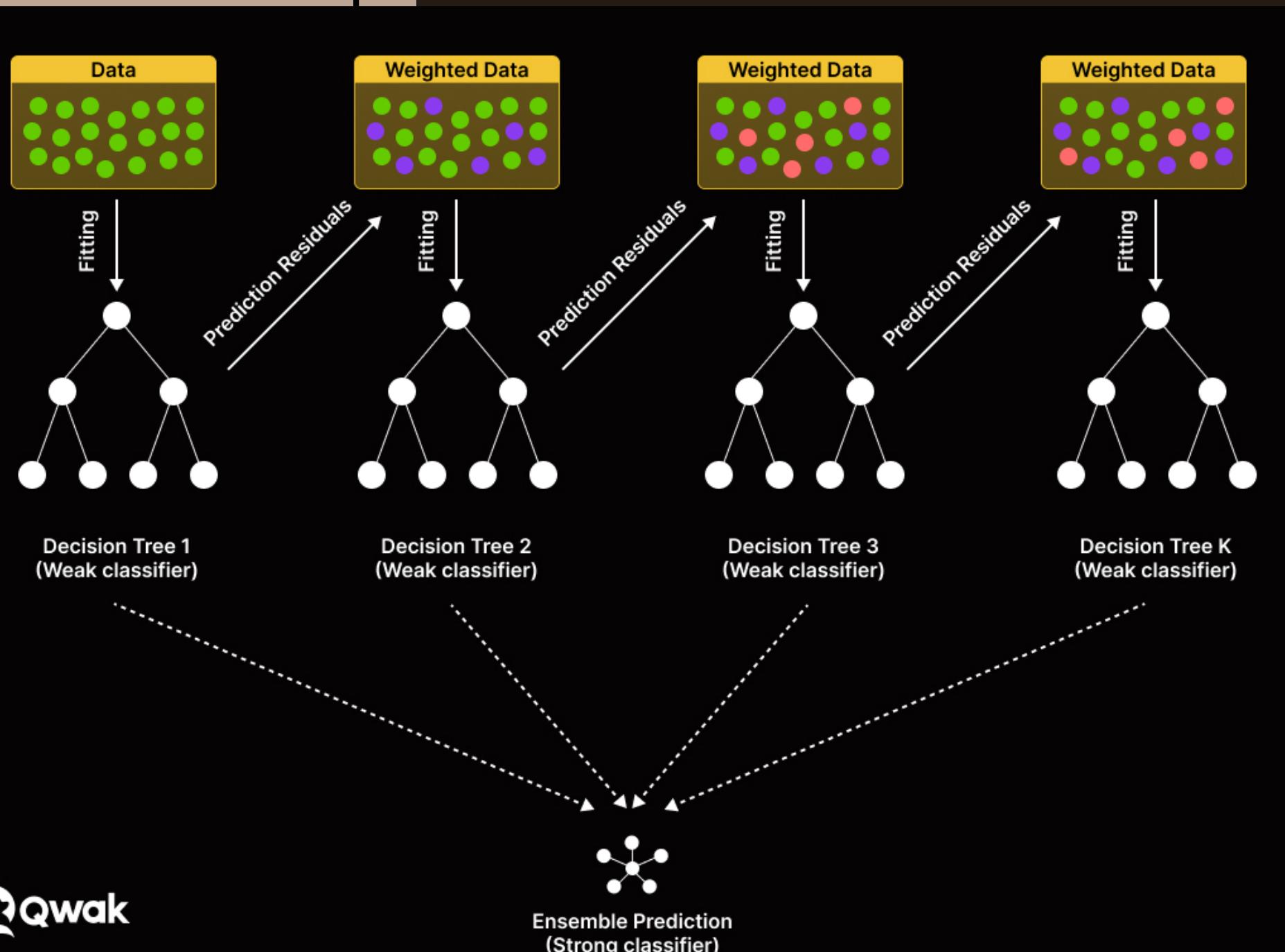
DASHBOARD DEVELOPMENT



PHASE 1

LITERATURE REVIEW

XGBOOST

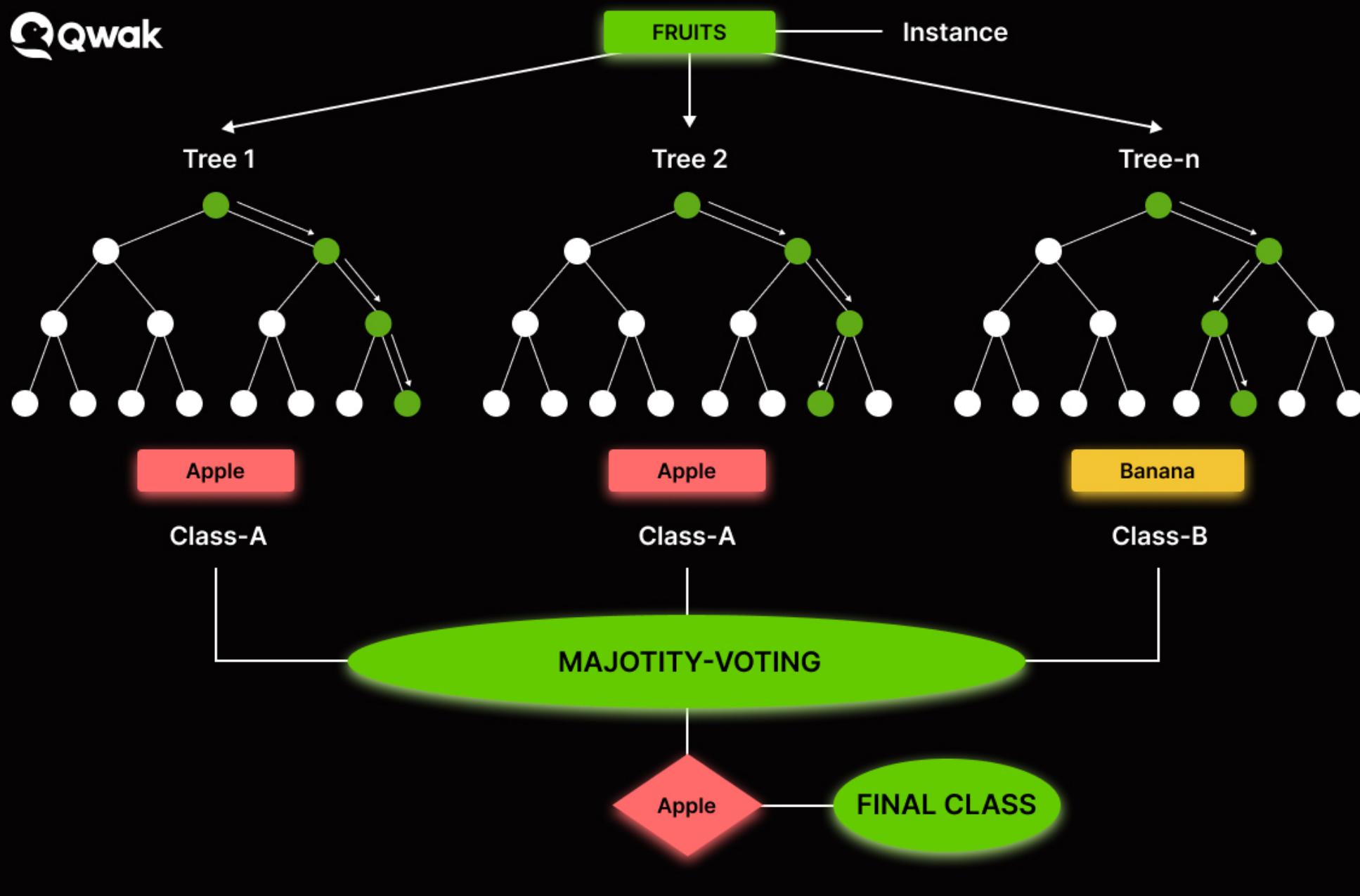


Qwak

XGBoost Model Architecture

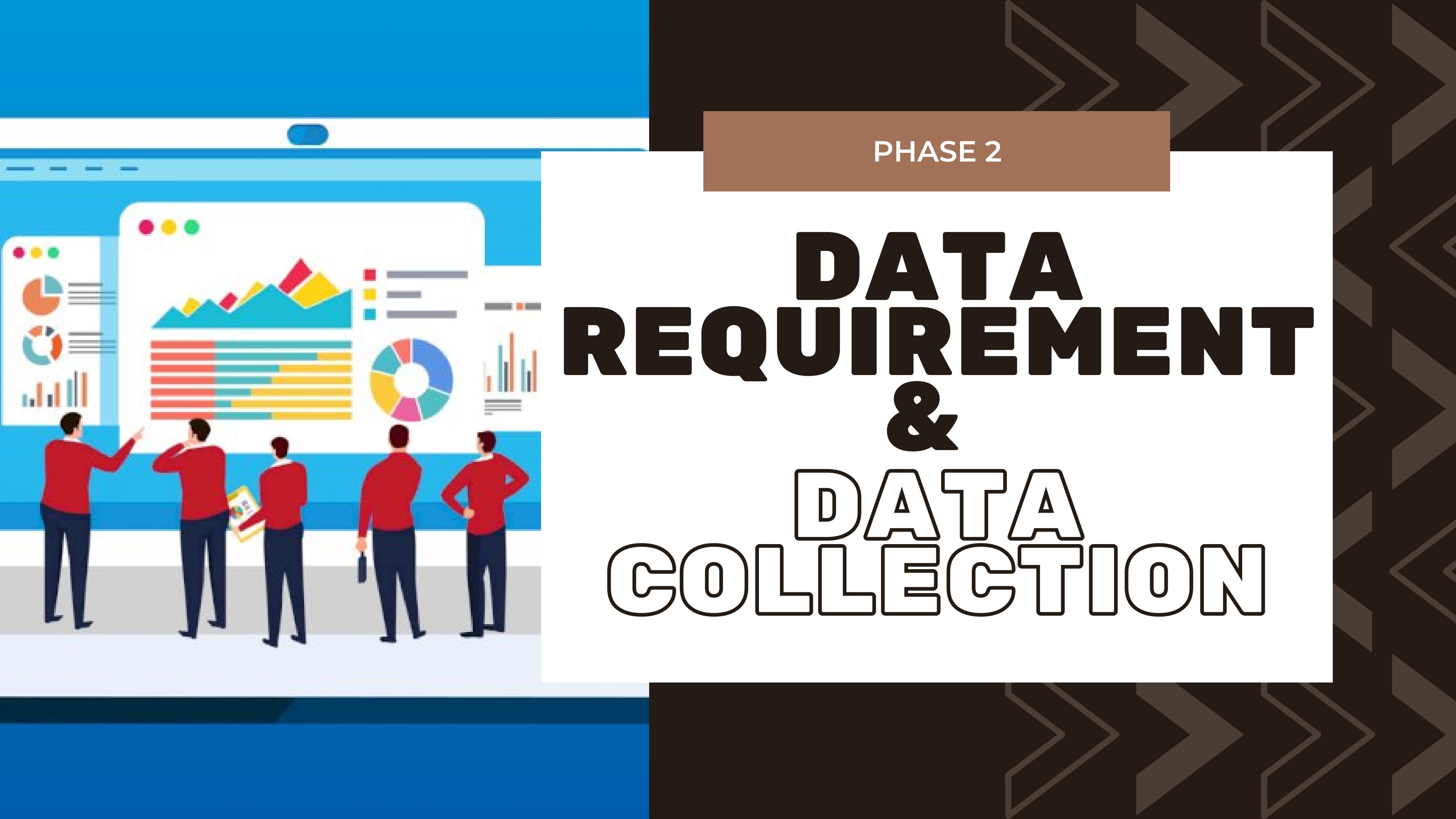
- XGBoost is a popular machine learning algorithm used for classification tasks.
- Combines multiple weak learners to create a strong model through boosting.
- Handles missing data effectively using surrogate splits.
- Optimized for parallel computing, making it fast and efficient.
- Outperforms other 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

The screenshot shows the homepage of the myMETdata website. At the top left is the Malaysian coat of arms and the METMalaysia logo. The top right features navigation links: Home, Product (with a dropdown arrow), Information (with a dropdown arrow), EN (with a dropdown arrow), Login, and a shopping cart icon with a red '0' notification. Below the header, the title 'Climate Sub-Categories' is centered. Eight blue square icons are arranged in two rows of four. The first row contains: Rainfall (cloud with rain), Temperature (thermometer), Relative Humidity (water droplet with percentage sign), and Wind (anemometer). The second row contains: Solar Radiation (sun with wavy arrows), Sunshine (sun with rays), Evaporation (cloud over water with upward arrows), and Cloud (cloud). Each icon has its corresponding label below it.

- Rainfall
- Temperature
- Relative Humidity
- Wind
- Solar Radiation
- Sunshine
- Evaporation
- Cloud

SOURCE : MET MALAYSIA

Products In Climate

Show 50 entries

Search:

No	Product Name	Types	File Format	Price
1	Daily Temperature at 2.00 p.m.	Daily	CSV	RM20
2	Daily Mean Temperature	Daily	CSV	RM20
3	Daily Minimum Temperature	Daily	CSV	RM20
4	Daily Mean Wet Bulb Temperature	Daily	CSV	RM20
5	Daily Maximum Wet Bulb Temperature	Daily	CSV	RM20
6	Daily Minimum Wet Bulb Temperature	Daily	CSV	RM20
7	Hourly Dry Bulb Temperature	Hourly	CSV	RM20
8	Monthly Wet Bulb Temperature	Monthly	CSV	RM30
9	Monthly Temperature	Monthly	CSV	RM30
10	Monthly Dry Bulb Temperature at 2.00 p.m.	Monthly	CSV	RM30
11	Monthly Dry Bulb Temperature at 8.00 a.m.	Monthly	CSV	RM30
12	Monthly Mean Maximum Temperature	Monthly	CSV	RM30
13	Monthly Highest Maximum Temperature	Monthly	CSV	RM30
14	Monthly Lowest Minimum Temperature	Monthly	CSV	RM30

TEMPERATURE DATA

Met Data > Iklim

Products In Climate

Show 50 entries

Search:

No	Product Name	Types	File Format	Price
1	Daily Maximum Surface Wind	Daily	CSV	RM20
2	Daily Mean Surface Wind Speed	Daily	CSV	RM20
3	Hourly Surface Wind	Hourly	CSV-PDF	RM20
4	Monthly Maximum Surface Wind	Monthly	CSV	RM30
5	Monthly Mean Surface Wind Speed	Monthly	CSV	RM30
6	Annual Wind Rose Summary	Summary	CSV-JPG	RM60
7	Seasonal Wind Rose Summary	Summary	CSV-JPG	RM60

Showing 1 to 7 of 7 entries

Previous 1 Next

WIND DATA


 Home Product ▾ Information ▾ EN ▾ Login 0

Met Data > Iklim

Products In Climate

No	Product Name	Types	File Format	Price
1	Daily Global Radiation	Daily	CSV	RM20
2	Hourly Global Radiation	Hourly	CSV	RM20
3	Hourly Cloud Cover	Hourly	CSV	RM20
4	Monthly Mean Daily Global Radiation	Monthly	CSV	RM30

Show 50 entries Search:

Showing 1 to 4 of 4 entries Previous 1 Next


 Home Product ▾ Information ▾ EN ▾ Login 0

Met Data > Iklim

Products In Climate

No	Product Name	Types	File Format	Price
1	Daily Mean Relative Humidity	Daily	CSV	RM20
2	Hourly Relative Humidity	Hourly	CSV	RM20
3	Monthly Mean Relative Humidity	Monthly	CSV	RM30
4	Monthly Mean Relative Humidity at 2.00 p.m.	Monthly	CSV	RM30
5	Monthly Mean Relative Humidity at 8.00 a.m.	Monthly	CSV	RM30
6	Monthly Mean Maximum Relative Humidity	Monthly	CSV	RM30
7	Monthly Mean Minimum Relative Humidity	Monthly	CSV	RM30

Show 50 entries Search:

Showing 1 to 7 of 7 entries Previous 1 Next

SOLAR RADIATION DATA

RELATIVE HUMIDITY DATA

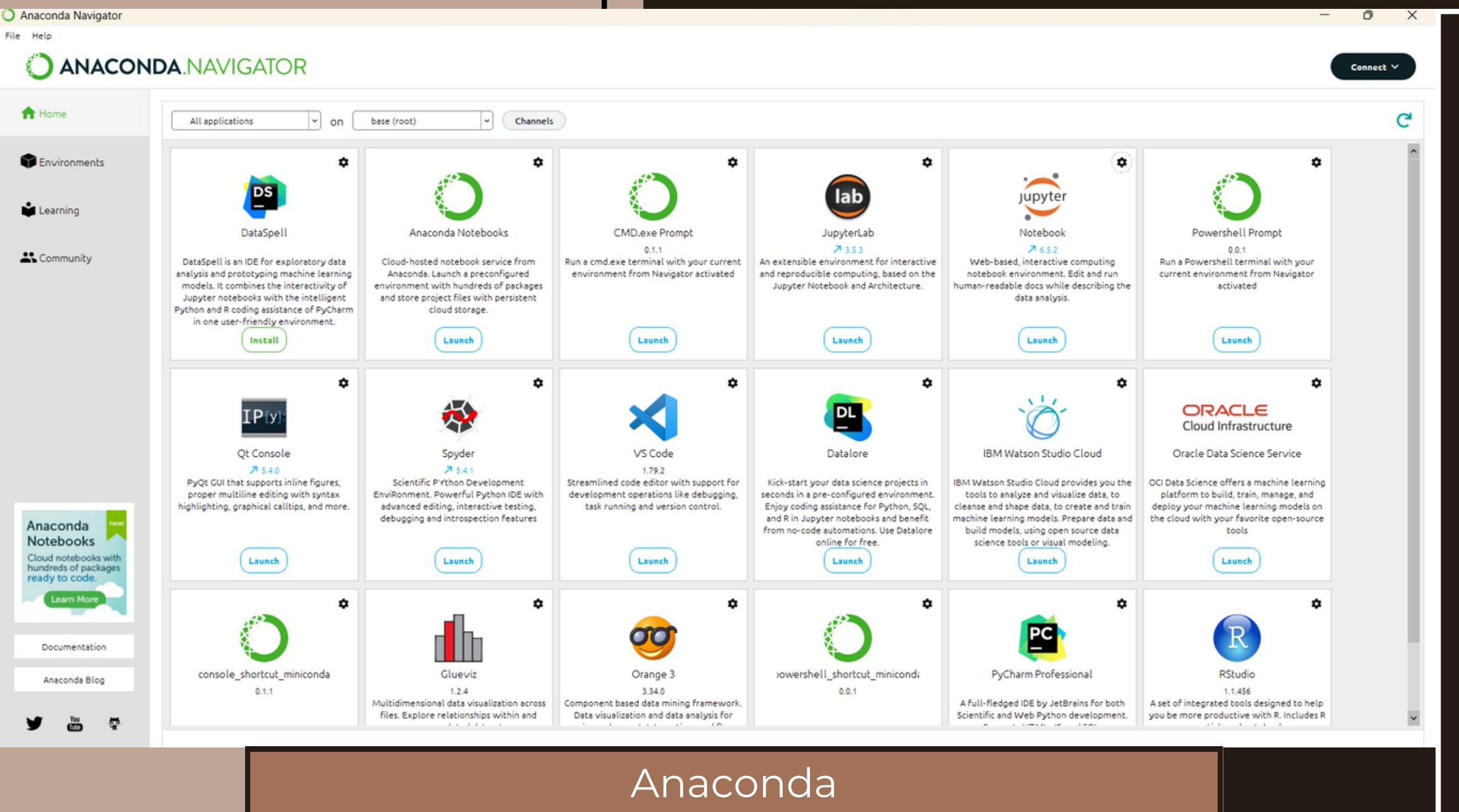


PHASE 2

MACHINE LEARNING

MODEL DEVELOPMENT

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

The screenshot shows a Jupyter Notebook interface with the title "Machine Learning Model". The notebook contains several code cells:

```
In [ ]: import pandas as pd
from sklearn.ensemble import RandomForestClassifier
from xgboost import XGBClassifier
from sklearn.model_selection import train_test_split

In [ ]: # Load and preprocess the dataset
df = pd.read_csv('microclimate_dataset.csv')

In [ ]: # Data preprocessing
# Drop irrelevant columns and rename columns (customize based on dataset)
cols_to_drop = ['irrelevant_column1', 'irrelevant_column2']
df = df.drop(columns=cols_to_drop).rename(columns={'old_column1': 'new_column1', 'old_column2': 'new_column2'})

In [ ]: # Split the data into training and testing sets
X = df.drop(columns='target_variable')
y = df['target_variable']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=8)

In [ ]: # Build a pipeline for Random Forest and XGBoost models
rf = RandomForestClassifier(random_state=8)
xgb = XGBClassifier(random_state=8)

In [ ]: # Evaluate the models
rf_accuracy = rf.score(X_test, y_test)
xgb_accuracy = xgb.score(X_test, y_test)

In [ ]: # Print the accuracy of each model
print("Random Forest Accuracy:", rf_accuracy)
print("XGBoost Accuracy:", xgb_accuracy)
```

Random Forest Model Architecture

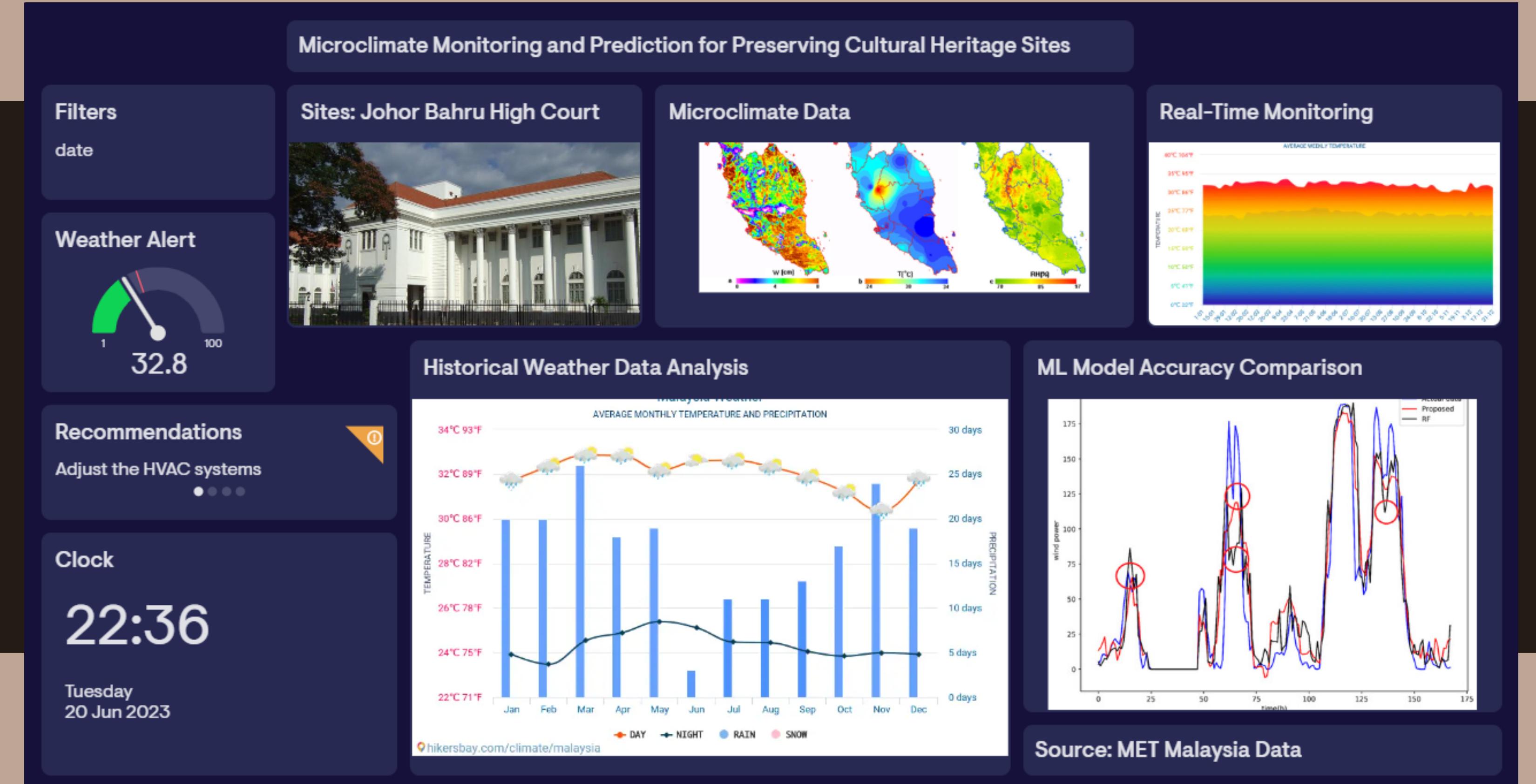
- Open-source web application that allows users to create and share documents containing live code, visualizations, and explanatory text.
- Provides an interactive computing environment, enabling users to write and execute code in small cells or blocks.
- Allows for the combination of code, text, and 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.

CS
BOARD



DASHBOARD DEVELOPMENT

DASHBOARD PROTOTYPE





Iman Aidi Elham | PSM1

**THANK
YOU**