

Presentation by **Iman Aidi Elham**

# **PRESERVING CULTURAL HERITAGE SITES THROUGH**

## **RANDOM FOREST AND XGBOOST ALGORITHM FOR MICROCLIMATE MONITORING AND PREDICTION**

Supervised by Assoc. Prof. Dr. Shahizan bin Othman

A20EC5006

Universiti Teknologi

Malaysia

2023

# RESEARCH OUTLINE

**Introduction**

**Problem Background**

**Objectives**

**Scopes**

**Methodology**

**Literature Review**

**Research  
Workflow**

**Research  
Significance**

# 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

- This country 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 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

# Scopes

## Scope 1

— X

The 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, wind speed, rainfall and solar radiation.

## 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 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 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.

# METHODOLOGY

**Phase 1**

**LITERATURE REVIEW**

**Phase 2**

**DATA REQUIREMENT & DATA  
COLLECTION**

**Phase 3**

**MACHINE LEARNING MODEL  
DEVELOPMENT**

**Phase 4**

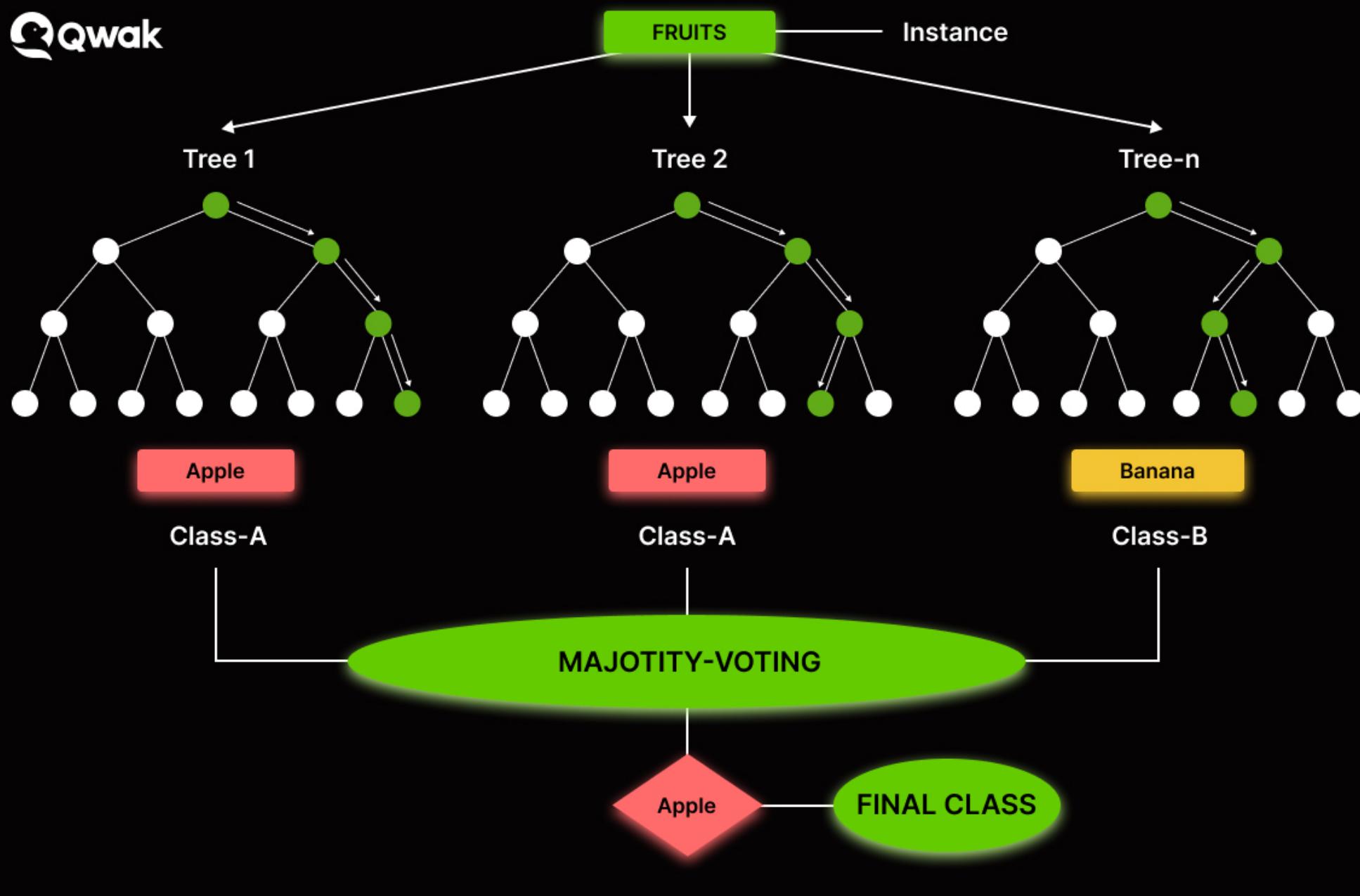
**DASHBOARD DEVELOPMENT**



PHASE 1

# LITERATURE REVIEW

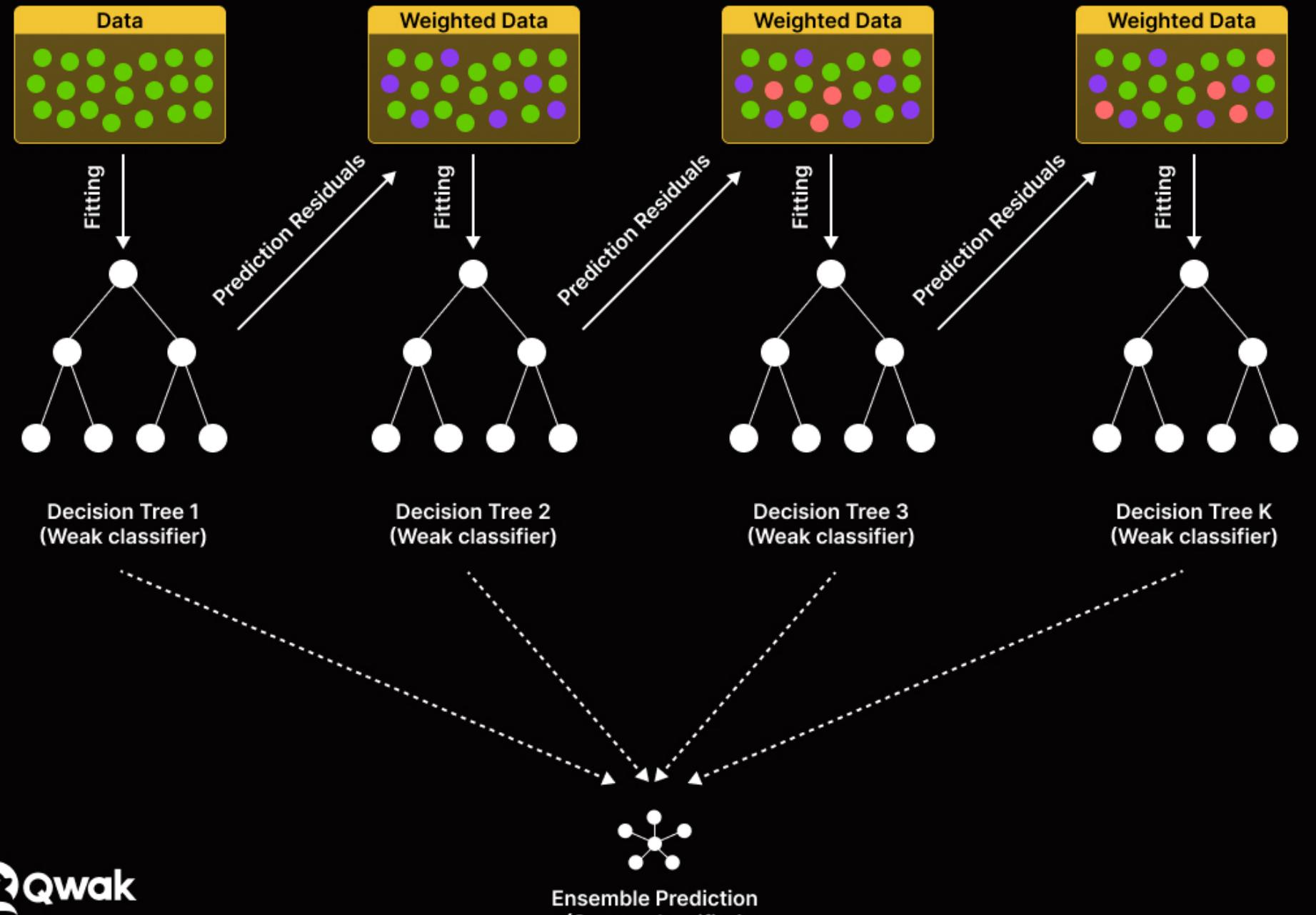
# RANDOM FOREST



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.

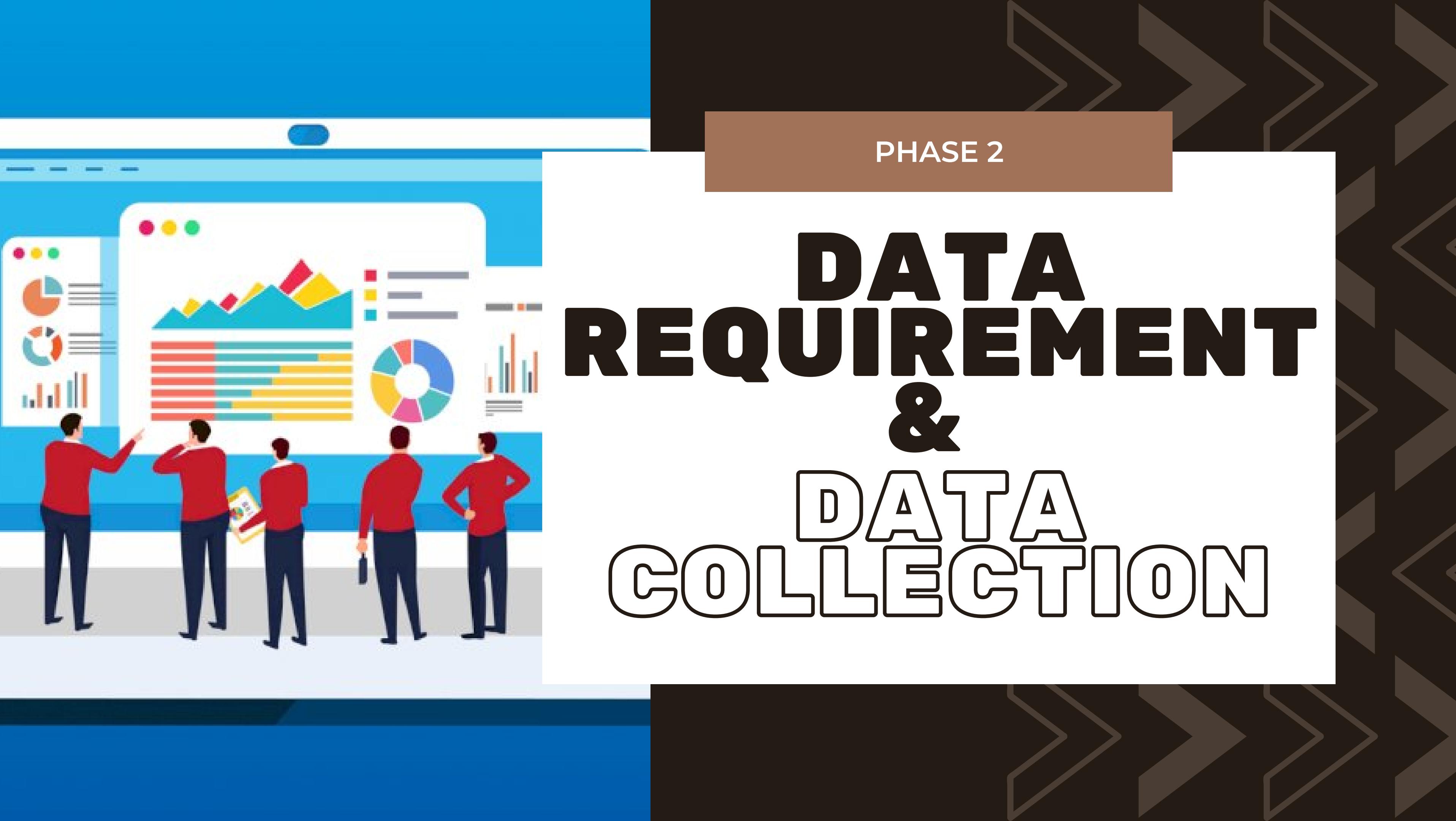
# XGBOOST



Qwak

XGBoost Model Architecture

- XGBoost is a popular machine learning algorithm used for classification and regression 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.



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'.

The main content area is titled "Climate Sub-Categories" and displays eight icons arranged in two rows of four:

- Rainfall:** An icon of a white cloud with three yellow raindrops falling against a blue background.
- Temperature:** An icon of a white thermometer with a blue scale and a blue bulb against a blue background.
- Relative Humidity:** An icon of a large blue water droplet with a white percentage sign (%) inside against a blue background.
- Wind:** An icon of a wind vane with three blades and a blue cone at the base against a blue background.

- Solar Radiation:** An icon of a sun with rays and a blue wavy line representing solar energy against a blue background.
- Sunshine:** An icon of a sun with rays against a blue background.
- Evaporation:** An icon of a white cloud with three blue arrows pointing upwards from a blue wavy line representing water against a blue background.
- Cloud:** An icon of a white cloud against a blue background.

**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**

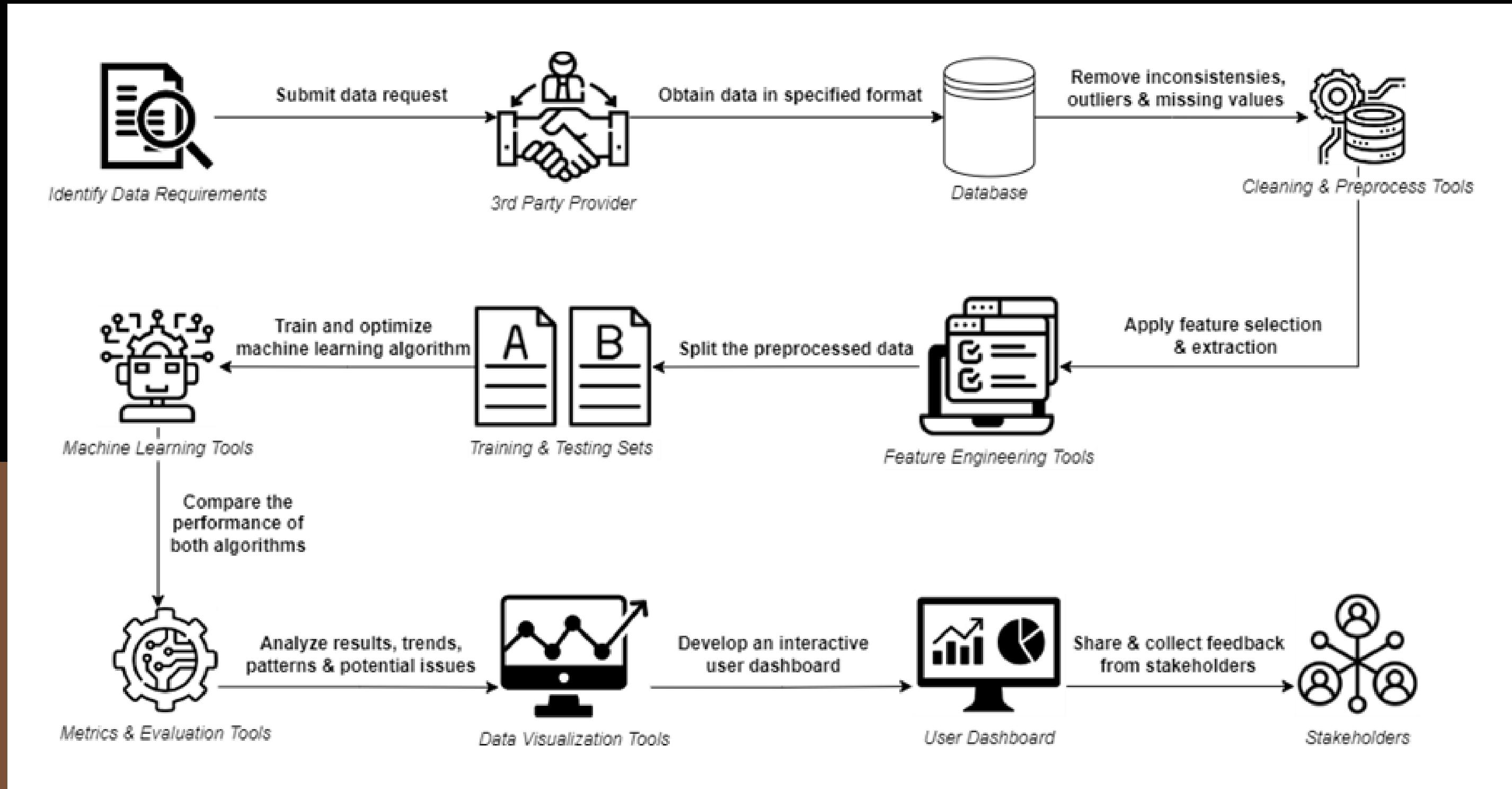
**RELATIVE HUMIDITY DATA**



PHASE 2

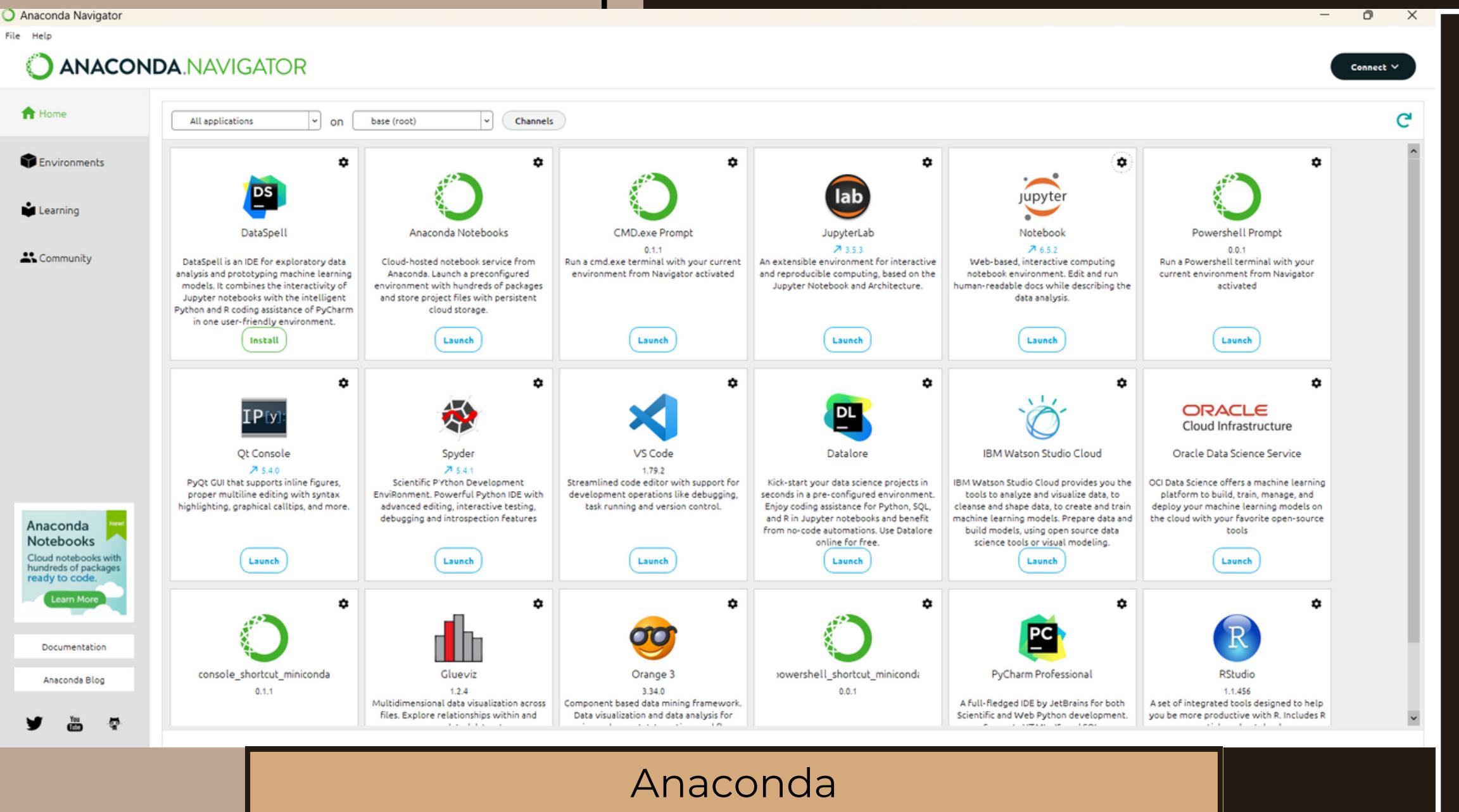
# MACHINE LEARNING

# MODEL DEVELOPMENT



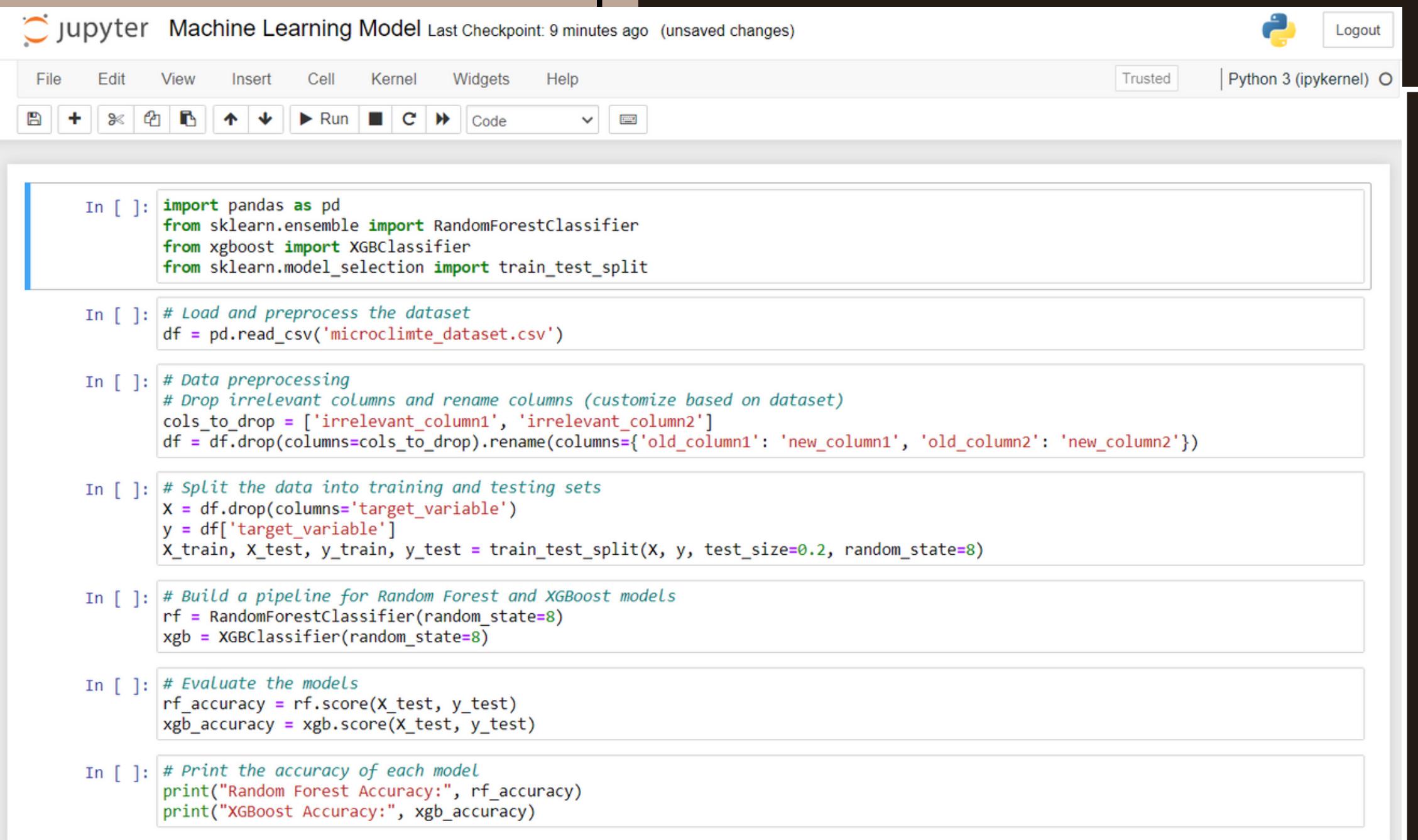
# MACHINE LEARNING WORKFLOW

# 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 following 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)
```

Jupyter Notebook

- 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.

# LIBRARIES & MODULES



CS  
BOARD



# DASHBOARD PROTOTYPE

JOHOR BAHRU HIGH COURT 1.4560° N, 103.7602° E Mon 03 July @ 10: 40: 04 Go to Map / Change Site

**Weather Station (MET Malaysia)**

Temperature <b>33.5°C</b> 29°C 35°C	Wind Speed <b>10.0 km/h</b>	Solar Radiation <b>1.0 kWh/m²</b>
Humidity <b>77%</b>	Rainfall <b>9.5 mm</b>	Johor Bahru High Court

**Maintenance Status**

Drainage	Good Service
Ceilings	Good Service
Floors	Good Service
HVAC	Need Adjustment
Rooftop	Warning
Walls	Good Service

**Historical Weather Data Analysis**

**ML Accuracy Comparison**

**Real-Time Monitoring**

**Ozon (O<sub>3</sub>)** Good 18 60.47 µg/m³

**NO<sub>2</sub>** Moderate 54 90.87 µg/m³

**SO<sub>2</sub>** Good 12 20.20 µg/m³

**PM<sub>2.5</sub>** Good 07 10.75 µg/m³

**PM<sub>10</sub>** Good 19 12.57 µg/m³

**CO** Unhealthy 70 12.57 µg/m³

**Weather Heat Map**

**Calendar**

SUN	MON	TUE	WED	THU	FRI	SAT
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30					

# DASHBOARD PROTOTYPE

**SULTAN IBRAHIM BUILDING** 1.4560° N, 103.7602° E Mon 03 July @ 10:30:01 [Go to Map / Change City](#)

**Weather Station (MET Malaysia)**

Temperature <b>32.0°C</b>	Wind Speed <b>10.0 km/h</b>	Solar Radiation <b>1.0 kWh/m²</b>
Humidity <b>77%</b>	Rainfall <b>9.5 mm</b>	

**Sultan Ibrahim Building**

**Forecast (MET Malaysia)**

Thu, July 06 <b>Sunny</b>	<b>31°C</b>
	<b>35°C</b>
Fri, July 07 <b>Sunny</b>	<b>31°C</b>
	<b>35°C</b>
Sat, July 08 <b>Cloudy</b>	<b>28°C</b>
	<b>30°C</b>

**Maintenance Status**

Drainage	Good Service
Ceilings	Good Service
Floors	Good Service
HVAC	Need Adjustment
Rooftop	Warning
Walls	Good Service

**Historical Weather Data Analysis**

**ML Accuracy Comparison**

**Real-Time Monitoring**

**Weather Heat Map**

**Real-Time Monitoring (Johor Bahru)**

Ozon (O <sub>3</sub> ) <b>Good</b>	<b>18</b>
NO <sub>2</sub> <b>Moderate</b>	<b>54</b>
SO <sub>2</sub> <b>Good</b>	<b>12</b>
PM <sub>2.5</sub> <b>Good</b>	<b>07</b>
PM <sub>10</sub> <b>Good</b>	<b>19</b>
CO <b>Unhealthy</b>	<b>70</b>

**Calendar**

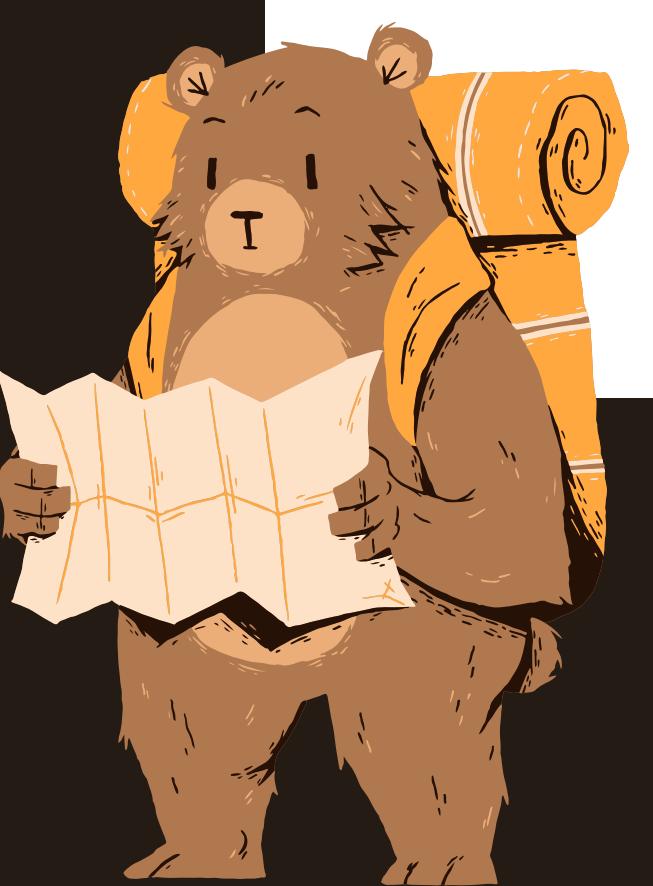
January 2023

SUN	MON	TUE	WED	THU	FRI	SAT
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30					

# SIGNIFICANCE OF RESEARCH

This study has essential significance for the preservation and conservation of natural resources and heritage sites in Johor Bahru, Malaysia. The findings of this study are expected to have practical applications for professionals involved in land use planning, landscape management, archaeological preservation, and public administration, as they strive to effectively manage cultural heritage sites and promote environmental sustainability through evidence-based strategies.

The study offers valuable insights and technical guidance regarding the selection of appropriate machine learning algorithms, proper interpretation, and preventive maintenance strategies. These findings can serve as a foundation for future research endeavors and decision-making processes in the field of microclimate monitoring and prediction for cultural heritage sites. By understanding the strengths and limitations of different machine learning algorithms, researchers and practitioners can make informed choices when selecting the most suitable models for their specific contexts.



**THANK  
YOU**

<https://youtu.be/hIkXnx3eUYY>