



**UTM**  
UNIVERSITI TEKNOLOGI MALAYSIA

**SCHOOL OF COMPUTING**  
Faculty of Engineering

Project Proposal Form  
MCST1043 Sem: 2  
Session: 2024/25

## SECTION A: Project Information.

---

Program Name: **Masters of Science (Data Science)**  
Subject Name: **Project (MCST1043)**  
**1**  
Student Name: NURHAFIZAH BINTI MOHD YUNOS  
Metric Number: MCS241048  
Student Email & Phone: [nurhafizah99@graduate.utm.my](mailto:nurhafizah99@graduate.utm.my) & 0143680885  
Project Title: Forecasting Malaysian Rice Production Using Historical Climate Data and Machine Learning Algorithms  
Supervisor 1: .....  
Supervisor 2 / .....  
Industry .....  
Advisor(if any): .....

---

## SECTION B: Project Proposal

### Introduction:

Rice is a staple food and a crucial agricultural commodity in Malaysia, playing a vital role in ensuring national food security. However, rice production is highly influenced by climatic conditions such as rainfall, temperature, humidity, and solar radiation. Accurate forecasting of rice production is essential for effective policy-making, resource allocation, and sustainable agricultural planning. Traditional forecasting methods often fail to capture the complex, non-linear relationships between climate variables and agricultural output.

This study proposes the use of machine learning techniques—Random Forest Regressor, Support Vector Regression (SVR), and Long Short-Term Memory (LSTM)—to build predictive models that forecast Malaysian rice production using historical climate data. The project aims to explore how these models perform in capturing temporal patterns and predicting future trends, providing valuable insights for decision-makers in the agriculture sector.

**Problem Background:**

Climate change has introduced significant variability into agricultural systems, affecting crop yields and threatening food security. In Malaysia, paddy (rice) production is particularly sensitive to changes in weather patterns. Despite its importance, there is limited application of advanced machine learning models in forecasting rice production at both regional and national levels. Existing forecasting tools lack the capability to incorporate high-resolution climate data and model complex interactions effectively.

**Problem Statement:**

There is currently a gap in accurate, scalable, and data-driven solutions for forecasting rice production in Malaysia using integrated climate datasets. Traditional statistical methods are insufficient for modelling the dynamic relationship between climatic factors and paddy output. This project addresses this gap by developing and evaluating machine learning-based forecasting models that utilise historical climate and agricultural data to improve prediction accuracy.

**Aim of the Project:**

The main aim of this project is to develop and evaluate machine learning models for forecasting Malaysian rice production using historical climate data. These models will support informed decision-making in agricultural planning, enhance food security strategies, and contribute to the growing field of smart agriculture in tropical regions.

**Objectives of the Project:**

The objectives of the project are:

1. To collect and preprocess historical data on rice production and climate variables across Malaysian states.
2. To conduct exploratory data analysis (EDA) to identify key drivers of rice production and understand seasonal and regional patterns.
3. To build predictive models using Random Forest Regressor, Support Vector Regression (SVR), and Long Short-Term Memory (LSTM) networks.
4. To evaluate and compare the performance of different models using metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and  $R^2$  score.
5. To provide recommendations for improving forecasting accuracy and supporting policy decisions related to food security and agricultural planning.

**Scope of the Project:**

1. Focus on monthly-level rice production data and associated climate variables from multiple Malaysian states.
2. Integration of data from official sources, including the Department of Statistics Malaysia (DOSM) and NASA POWER.
3. Application of machine learning algorithms for regression and time-series forecasting.
4. Development of a prototype forecasting framework with potential for real-time deployment.
5. Analysis of model interpretability and feature importance to guide future interventions.

**Expected Contribution of the Project:**

1. Accurate Forecasting Models: Development of models using Random Forest, SVR, and LSTM to predict rice production with high accuracy.
2. Data Integration Framework Combines agricultural and climate datasets at a state-month level, enabling detailed regional analysis.
3. Policy Support Tool: Provides insights for better agricultural planning, resource

allocation, and food security strategies.

4. Feature Engineering Insights: Highlights important predictors like planted area, lagged yield, and climatic factors influencing production.
5. Foundation for Future Research: Opens opportunities for incorporating soil quality, irrigation, and socio-economic factors into forecasting models.

**Project Requirements:**

Software: Python (Jupyter Notebook, Google Colab), Pandas , NumPy , SciTensorFlow/Keras, Matplotlib , Seaborn , Plotly, StandardScaler , Train, Model Evaluation Metrics

---

Hardware: Computer with a minimum of 8GB RAM  
GPU recommended for deep learning models (e.g., LSTM)

---

Technology/Technique/Methodology/Algorithm: Data Collection: From DOSM and NASA POWER datasets  
Data Preprocessing: Handling missing values, normalisation, time-based engineering  
Exploratory Data Analysis (EDA): Correlation analysis, trend visualisation, pattern exploration  
Model Development :  
1. Random Forest Regressor: For capturing non-linear relationships and feature importance  
2. Support Vector Regression (SVR) For robustness in handling scalar data  
3. Long Short-Term Memory (LSTM): For modelling sequential and temporal dependencies  
Model Evaluation: MAE, RMSE, R<sup>2</sup> Score

**Type of Project (Focusing on Data Science):**

☐ Data Preparation and Modelling

☐ Data Analysis and Visualisation

☐ Business Intelligence and Analytics

☐ Machine Learning and Prediction

☐ Data Science Application in the Business Domain

**Status of Project:**

☐ New

☐ Continue

If continued, what is the previous title? \_\_\_\_\_

---

**SECTION C: Declaration****I declare that this project is proposed by:**

☐ Myself

☐ Supervisor/Industry Advisor (\_\_\_\_\_)

Student NURHAFIZAH BINTI MOHD YUNOS

Name: .....

*Nurhafizah*

Signature

17/4/2025

Date

---

#### SECTION D: Supervisor Acknowledgement

The Supervisor(s) shall complete this section.

**I/We agree to become the supervisor(s) for this student under the aforesaid proposed title.**

Name of Supervisor 1: .....

Signature

Date

Name of Supervisor 2 (if any): .....

Signature

Date

---

#### SECTION E: Evaluation Panel Approval

The Evaluator(s) shall complete this section.

**Result:**

☐ FULL APPROVAL

☐ CONDITIONAL  
APPROVAL (Major)\*

☐ CONDITIONAL APPROVAL  
(Minor)

☐ FAIL\*

\* The student has to submit a new proposal form considering the evaluators' comments.



.....

.....

.....

.....

.....

.....

.....

.....

.....

Name of Evaluator 1: \_\_\_\_\_

.....

**Signature**

**Date**

Name of Evaluator 2: \_\_\_\_\_

.....

**Signature**

**Date**