FORECASTING MALAYSIAN RICE PRODUCTION USING HISTORICAL CLIMATE DATA AND MACHINE LEARNING ALGORITHMS

NURHAFIZAH BINTI MOHD YUNOS

UNIVERSITI TEKNOLOGI MALAYSIA

TABLE OF CONTENTS

TITLE **PAGE**

CHAPTER 1 INTRODUCTION...... 1

1.1 Problem Background...... 1

1.2 Problem Background...... 2

1.3 Problem Statement...... 3

1.4 Research Gaps...... 4

1.5 Significance of the Study...... 5

1.6 Scope and Limitations...... 5

1.7 Organisation of the Thesis...... 7

1.8 Summary...... 7

CHAPTER 1

INTRODUCTION

1.1 Problem Background

Rice is a vital part of Malaysia's food system. It's not just a staple on dinner tables across the country—it also plays a key role in national food security and supports the livelihoods of many rural communities. Although the government has made various efforts to boost local production, Malaysia still depends on rice imports to meet the needs of its population. According to the Department of Agriculture (DOA) Malaysia, rice makes a notable contribution to the agricultural sector's GDP, highlighting the importance of maintaining consistent and sustainable production.

But growing rice in Malaysia isn't without its challenges. Farmers are dealing with issues like unpredictable weather, changes in land use, limited water supply, and pest outbreaks. Climate change is making things even more uncertain, bringing shifts in rainfall, rising temperatures, and more extreme weather like floods and droughts. All of this directly affects how rice grows, how much can be harvested, and the overall stability of the farming cycle.

In response, there has been increasing interest in using data to better plan and manage agricultural activities. If we can predict rice yields more accurately, it would help farmers, planners, and policymakers stay ahead of potential shortages, manage resources wisely, and develop strategies to deal with climate-related risks. However, traditional forecasting methods often fall short—they usually depend on statistical models that might not fully capture the complex relationship between climate and crop performance.

That's where machine learning (ML) comes in. As a branch of artificial intelligence, ML is well-suited for handling complex patterns and large amounts of data. By combining past climate data with machine learning algorithms, we can create models that offer more accurate and adaptable forecasts, tailored to local farming conditions. This study looks into how ML can be used to predict rice production in Malaysia, to support a more resilient, data-driven approach to agriculture.

1.2 Problem Background

Accurately predicting rice production is crucial, not just for maintaining food security but also for helping policymakers make informed decisions. Unfortunately, the forecasting methods currently used in Malaysia often fall short. They tend to overlook important climate details and rely on models that assume simple, linear relationships, which don't fully reflect the complex and changing environmental factors that influence crop yields.

There’s still a shortage of research focused on using machine learning specifically to forecast rice production in Malaysia. Few studies have tapped into high-resolution historical climate data or explored how this information can be combined with advanced models to improve prediction accuracy. By integrating this kind of detailed data with machine learning, we could provide much more reliable forecasts, giving farmers, planners, and policymakers the insights they need to plan effectively.

This research aims to tackle several key challenges:

1. The limited use of climate data in current rice yield forecasting models.
2. The lack of localised machine learning models that reflect Malaysia’s unique rice-growing conditions.
3. The a need for deeper analysis of how different climate factors impact rice yields over time.

1.3 Problem Statement

This study sets out to achieve the following goals:

a) Examine historical data on rice production and climate trends in Malaysia.

b) Identify which climate factors have the strongest impact on rice yields.

c) Build predictive models using selected machine learning techniques.

d) Assess and compare how well different machine learning models perform in forecasting rice production.

1.4 Research Gaps

To meet these objectives, the study aims to answer several key questions:

a) Which climate variables most significantly influence rice production in Malaysia?

b) Which machine learning model delivers the most accurate rice yield forecasts?

c) How well do these models perform when tested against real-world rice production data?

1.5 Significance of the Study

This research has practical importance for various groups in the agriculture sector:

a) Policymakers: Reliable forecasting can support the development of data-driven policies around food security, import strategies, and resource planning.

b) Agricultural Planners: Accurate predictions can help with scheduling planting, irrigation, and harvesting activities.

c) Farmers: Better forecasts allow for smarter decisions about managing crops, using inputs, and timing their sales.

d) Researchers: This work adds to the growing research on how machine learning can be used in agriculture, especially in tropical regions like Malaysia.

1.6 Scope and Limitations

This study focuses on predicting rice production in Malaysia using national-level historical data over a specific time frame. It includes key climate variables such as temperature, rainfall, humidity, solar radiation, and wind speed, sourced from official databases like the Department of Statistics Malaysia (DOSM) and NASA POWER.

However, there are a few limitations to note:

a) The accuracy of the models depends heavily on the availability and quality of historical data.

b) The models may not capture the full influence of socio-economic or farm-level practices on rice yields.

c) The findings are specific to rice production in Malaysia and may not directly apply to other crops or regions without adjustments.

1.7 Organisation of the Thesis

The structure of this thesis is as follows:

a) Chapter 1: Introduction – Outlines the background, research problem, objectives, questions, significance, scope, and the overall structure of the thesis.

b) Chapter 2: Literature Review – Explores existing research on rice farming in Malaysia, the impact of climate change on agriculture, current forecasting methods, and the use of machine learning in crop prediction.

c) Chapter 3: Methodology – Details the research approach, data sources, preprocessing steps, chosen machine learning techniques, and evaluation methods.

d) Chapter 4: Initial Findings and Analysis – Shares the model results, discusses the outcomes, and interprets the insights gained.

e) Chapter 5: Conclusion, Recommendations, and Future Work – Summarises the key contributions, outlines practical implications, notes the study's limitations, and suggests directions for future research.

1.8 Summary

This thesis investigates how machine‑learning (ML) techniques, fueled by high‑resolution historical climate data, can improve the accuracy of rice‑yield forecasts in Malaysia—an essential step toward bolstering national food security and farmer resilience. After outlining the country's heavy reliance on rice, the study highlights key production challenges—unpredictable weather, limited water, land‑use shifts and pests—made worse by climate change. Existing Malaysian forecasting tools rarely integrate detailed climate information or capture the non‑linear dynamics between weather and crop performance, leaving a clear research gap. Accordingly, the work (i) identifies the climate variables that most strongly drive rice yields, (ii) develops and compares several ML models for prediction, and (iii) tests each model against real‑world production data to determine the most reliable approach. The outcomes are intended to guide policymakers in food‑security planning, help agricultural planners and farmers optimise field operations, and extend academic understanding of data-driven agriculture in tropical settings. While the analysis draws on robust national datasets from DOSM and NASA POWER, results remain sensitive to data quality and do not fully capture socio-economic factors at farm level, limiting direct transferability to other crops or regions. The thesis is organised into five chapters: an introduction; a literature review of rice, climate impacts and forecasting methods; a detailed methodology; empirical findings; and conclusions with recommendations for future work.