Comprehensive Analysis of Rainfall Forecasting in Selangor Using Machine Learning Techniques

Muhammad Zamil Syafiq Bin Zanzali Supervisor: Dr. Tay Chai Jian

January 2025

# Abstract

**Background:** Accurate rainfall forecasting is critical for flood management, agricultural planning, and water resource management in Selangor, Malaysia, where rainfall patterns exhibit high variability due to tropical climate influences.

**Methods:** This study employs six machine learning models—Artificial Neural Networks (ANN), Mul- tiple Linear Regression (MLR), K-Nearest Neighbors (KNN), Random Forests (RF), Gradient Boosting (XGBoost), and ARIMA—to predict rainfall patterns. Meteorological data from 2012-2020, including tem- perature, humidity, and wind speed measurements, were processed using mean imputation for missing values and Min-Max normalization.

**Results:** Model performance was evaluated using MAE, MSE, RMSE, and R2 metrics. The Random

Forest model demonstrated superior performance with an R2 of 0.89 and RMSE of 2.7mm. Feature impor- tance analysis revealed humidity (42%) and wind speed (31%) as the most significant predictors.

**Conclusion:** Machine learning techniques, particularly ensemble methods like Random Forest, provide accurate rainfall forecasts that can be integrated into early warning systems. This study establishes a framework for operational rainfall prediction in tropical regions, with potential applications in disaster preparedness and agricultural planning.

# Introduction

## Research Background

Rainfall patterns in Selangor region of Malaysia fluctuate widely partially driven by the tropical climate. In Selangor precipitation patterns are significantly influenced by tropical climate with the heaviest rainfall happening between October and December. November is the peak of this season where 324 mm of rainfall is experienced across 28 days. In October 222 mm of rainfall is experienced while in December 246 mm of rainfall is experienced. At the beginning of the year the amount of rainfall is relatively lower. January and February receive 148 mm and 102 mm respectively. However, April receives a rainfall of 241 mm which is comparable to precipitation received in peak season. During the summer months of June and July relatively lower rainfall amounts of 145 mm and 135 mm respectively are received [[5].](#_bookmark4) These seasonal patterns have a major influence on local ecosystems as well as agriculture activities and water management. The Malaysian Meteorological Department [**?**] analysed annual rainfall data from 1951 to 2023 and found there has been an upward trend in the amount of rainfall received in the country. This points to climate change that can lead to higher temperatures, rising sea levels, often occurrence of extreme events such as floods, disruption of habitats and agricultural activities, and economic losses. These fluctuations make it difficult to accurately forecast climate patterns. Climatic events such as frequent and heavy rainfall can lead to crop failure, floods, and water contamination. Similarly, seasons such as the monsoon have a significant influence on rainfall and its distribution. The Department of Irrigation and Drainage Malaysia [**?**] describes monsoon rains as “typically of long duration with intermittent heavy bursts and the intensity can occasionally exceed several hundred mm in 24 hours”. This can lead to floods in urban areas and disruption of agricultural

activities. Accurate forecasting will help the Selangor State Government in mitigating the effects of these events. Equipped with accurate forecasts the state government can put in place well planned emergency as well as disaster and preparedness strategies.

Machine learning models have become a critical tool in analysis of meteorological data. When comparing machine learning models with conventional Numerical Weather Prediction models, it has been observed ma- chine learning models are superior at detecting intricate numerical and non-linear patterns in data [[2].](#_bookmark1) This makes machine learning a suitable approach for predicting rainfall in a tropical region like Selangor. Large amounts of meteorological data can be analysed using machine learning techniques such as support vector machines (SVM), gradient boosting, and artificial neural networks (ANN) to provide accurate temporal esti- mations. These methods that will be discussed later, use historical data such as temperature, humidity, wind speed, and rainfall to provide accurate forecasts which were hitherto impossible using traditional techniques such as linear regression.

The problem is critical in places such as Selangor, where rainy conditions have not been accurately forecasting posing several difficulties. Hydrological functions enhanced by better rainfall predictions enable timely decisions in crop production, disaster management including floods and landslides, and water man- agement. Due to improved accuracy levels of predictions, stakeholders will be in a position to save structures from destruction, people from hunger as well as resources from wastage.

Recent advances in machine learning have expanded possibilities for improving rainfall forecasting. Ma- chine learning methods like support vector machines, gradient boosting, and artificial neural networks have shown great potential in capturing both temporal and spatial patterns of rainfall. These models are able to improve forecasts by continuously learning from new data. In Selangor, using machine learning techniques and local meteorological data presents an opportunity to develop a forecasting system that is highly accurate.

## Problem Statement

Climate change has received significant global attention due to disastrous events it can cause. Rainfall is a major meteorological factor that is influenced by climate change. In Malaysia, rainfall patterns have changed causing floods and droughts. Selangor is one the states that has been affected by these changes in rainfall patterns. Disastrous floods happened consecutively in the years 2006 to 2008 and in the years 2010 and 2011. The years 1997, 1998, and 2008 had catastrophic dry periods [[7].](#_bookmark6) Agricultural decisions and productivity are significantly influenced by environmental variables particularly the amount of water available and rainfall. In Selangor the influence of these variables is significant and a threat to agricultural productivity. High and low rainfall affects crops. Although it is possible to mitigate low rainfall through irrigation, high rainfall usually damages crops and results in low agricultural productivity. Mitigation measures such as changing crop cycles and combining crop cycles have not been adequate. To adequately solve these problems technological solutions are required [[1].](#_bookmark0)

One of the technological solutions that can be used is availing accurate rainfall predictions. However, due to irregular occurrence of rainfall in Timur Region Selangor accurate prediction is difficult. This situa- tion can harm farming, cause floods, and cause difficulties in water resources planning. Traditional models such as linear regression may not provide accurate precipitation forecast especially in the tropics because the atmospheric behaviour is not easy to predict. For example, Kassem et al. [[3]](#_bookmark2) reported artificial neural networks were superior to linear regression in predicting monthly rainfall in Northern Cyprus. That study showed artificial neural networks were better at capturing relationships in coordinates, meteorological vari- ables, and rainfall resulting in more accurate prediction compared to linear regression. Traditional models such as linear regression are weak at capturing complex relationships especially when they are non-linear. Compared to models such as support vector machines and artificial neural networks, linear regression mod- els are poor at handling non-linear relationships. Conversely, support vector machines and artificial neural networks are difficult to interpret, computationally costly, and require large amounts of data [**?**]. Modern meteorological research does not face the limitations of small datasets and limited computational power that were prevalent several decades ago. Meteorological instruments and IoT sensors have enabled accumulation of large datasets. This situation enables use of advanced machine learning models such as support vector machines and artificial neural networks in predicting rainfall. Specifically, in Selangor large volumes of me- teorological data are available. Therefore, these advanced machine learning models can be used to accurate predict rainfall patterns. Insights obtained will be useful in agricultural, infrastructure, public health, and

water management planning.

## Research Questions

The specific research questions that will be investigated in this study are:

1. What are the machine learning models that can be used for rainfall prediction in Selangor?
2. How does the performance of different machine learning models differ?
3. What is the best model in forecasting rainfall pattern in Selangor?

## Research Objectives

The broad objective of this study is to investigate the use of machine learning models in predicting rainfall in Selangor region of Malaysia. The specific objectives are:

1. To employ machine learning models that can be used for predicting rainfall in Selangor.
2. To estimate and assess the performance of different machine learning models using performance metrics such as mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and R-squared.
3. To identify the best model for forecasting rainfall patterns in Selangor by comparing performance metrics and selecting the model with highest accuracy.

## Research Scopes

This research deals with rainfall prediction for Selangor, Malaysia where the rainfall has irregular tropical pattern and significantly affects sectors such as water supply and flood control, agriculture. These problems will be addressed in this work by utilising and comparing a number of machine learning algorithms with support vector machines (SVM), gradient boosting, and artificial neural networks (ANN). These methods were chosen due to the possibility of the interpretation of which dependencies – both linear and nonlinear ones – are present in the data. In the present study, meteorological data from Sepang/KL International Airport is employed for data analysis where necessary climatic factors embracing average temperature, relative humidity, wind velocity, and precipitation for the years between 2012 and 2020 are utilised. This is to make certain that the data collected are accurate and reliable to increase the efficiency of data analysis after it has been fed into the system therefore data cleaning, normalization of data, handling of missing values and feature engineering will be undertaken.To fully assess predictive performance, the model will be evaluated using measures like the Coefficient of Determination (*R*2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE).

## Significance of Study

The research focus on using machine learning for rainfall forecasting in Selangor. Machine learning techniques utilize historical data to identify complex relationships, resulting in more precise and current forecasts. This study improves the scientific understanding of based on rainfall forecasting by evaluating how well different machine learning algorithms capture detailed tropical rainfall patterns. It represents a major breakthrough in environmental prediction and building resilience since it expands the use of machine learning for tropical weather forecasting and offers a structure that can be adjusted for different climates. The forecasting results could help the government in enhancing disaster readiness.

# Literature Review

## Introduction

In tropical regions such as Selangor in Malaysia where extreme events such as high or low rainfall happen; accurate rainfall forecasting is critical. When managers are provided with accurate predictions, they are better placed to put mitigation measures in place. These measures can help in management of disruptive events such as floods, agricultural crop failure, and disruptions in water supply. Machine learning has emerged as a powerful technique for analysing rainfall data, discovering patterns in meteorological data, and accurately predicting rainfall. This chapter presents an exhaustive review of existing literature on use of machine learning for forecasting rainfall. Specifically, the strengths and limitations of each study are evaluated to identify research gaps that can be addressed in this study and future studies.

## Challenges in Rainfall Forecasting

Numerous studies have well documented challenges faced when predicting rainfall. Kundu et al. [[4]](#_bookmark3) have discussed some of these challenges. These authors note the primary challenge is the wide variability in rainfall patterns. Other challenges are scarcity of relevant meteorological variables such as soil, humidity, wind and temperature parameters which are essential. When these variables are not available the accuracy of prediction models is severely affected. Other human activities such as deforestation can also negatively affect the accuracy of rainfall prediction models. Even when advanced methodologies are used accurate prediction of rainfall is challenging as large volumes of data and collaboration are required.

The National Oceanic and Atmospheric Administration [[8]](#_bookmark7) notes forecasting weather phenomena is a difficult skill that requires meticulous observation and analyzing large amounts of data. Weather phenomena can be characteristically thunderstorms covering large areas or a small area that can last for a few hours or several days. The phases involved in weather forecasting are observation, prediction, and dissemination of results.

Ray et al. [[6]](#_bookmark5) discuss various challenges faced in predicting rainfall driven by landfalling tropical cyclones in India. Rainfall from these tropical cyclones especially when approaching landfall varies widely and is usually asymmetric. This pattern is often caused by wind, speed, land surface, and moisture parameters. That study found that increase or decrease in intensity of a tropical storm as it approached the coastline during landfall can change the characteristics of rainfall over land.

Selangor is a typical tropical environment characterized by widely fluctuating rainfall patterns. This variation makes accurate rainfall prediction challenging. These challenges arise because rainfall patterns are influenced by intricate relationships among atmospheric factors like variations temperature, humidity, and windspeed. Rainfall predictions are usually obtained from large scale computerized simulations of weather systems. Use of traditional prediction methods like numerical weather prediction fails at capturing events that happen in isolated areas. Furthermore, this problem is severe in areas that have widely varying rainfall patterns such as Selangor. These models are further limited by their high cost and their lack of flexibility to adjust to changes in rainfall patterns in real time. Machine learning is a viable alternative for overcoming challenges faced by these traditional models. Particularly, machine learning models are suited to capturing complex and non-linear relationships that exist in meteorological data. With these capabilities machine learning models are an essential tool for discovering patterns that exist in historical meteorological data.

## Overview of Machine Learning Techniques for Rainfall Prediction

Machine learning models are well suited to capture non-linear relationships that are a common feature in meteorological variables like temperature, windspeed, humidity, and precipitation. This makes machine learning models a robust technique for analysing meteorological data. This section presents an exhaustive review of literature that has examined use of different machine learning models for rainfall prediction.

# Methodology

## Introduction

This chapter presents the steps that will be followed in identifying the machine learning algorithm that provides the highest accuracy in predicting rainfall in Selangor. The steps involved are exhaustive review of available literature, identifying the problem to be investigated, collecting relevant data, pre-processing data to assure its suitability, model training, tuning model parameters, and evaluating models. This structured approach will ensure all critical steps are followed. It is expected this approach will help in meeting study objectives.

## Research Design

This research design will act like a blueprint that will be followed in every stage of the study. The core objective is to compare machine learning algorithms and identify the algorithm that provides the highest prediction accuracy. A data driven approach is followed whereby historical weather data such as precipitation, temperature, humidity, and windspeed are the foundation of the study. A data science lifecycle that involves data gathering, pre-processing, parameter tuning, and model evaluation is followed.

## Data Science Methodology

### Literature Review

The first step in carrying out a study is reviewing available literature. Extant literature on machine learning models used for predicting rainfall was reviewed. From reviewed literature it was evident machine learning is an established technique in rainfall forecasting. Reviewed literature revealed machine learning models are primarily used for forecasting the amount of rainfall or classifying rainfall to several categories such as rain/no rain or intensity of rainfall such as low/medium/high. To a lesser extent machine learning were also used to identify critical factors that affect rainfall. Commonly used machine learning methods were support vector machines, decision trees, K-nearest neigbour, logistic regression, gradient boosting, XGBOOST, linear regression, and artificial neural networks. With the exception of logistic regression all the other machine learning models can be used to predict a quantitative amount of rainfall. It was evident in almost all studies a train and test subset were used. This provides a subset for training the model and another subset not used for training that will be used to evaluate model performance. Reviewed literature showed data preprocessing steps such as checking missing values, imputing missing values, checking out of range values, and normalizing quantitative variables to a common range are critical to performance of a machine learning model. From the literature it was observed that some machine learning models have hyperparameters that need to be tuned to achieve high prediction accuracies. These principles that are well established in the literature will be incorporated in this study.

### Problem Identification

Climate change has resulted in disruption of established weather patterns. This is a global phenomenon that can lead to extreme rainfall events such as too little or too much rainfall. These events have significant impact on public health, infrastructure, and agriculture. Although economic activities in Selangor are not primarily agricultural, extreme rainfall events need proper planning and mitigation. As a largely urbanized area, flooding from extreme rainfall events such as too much rainfall can cause major disruptions in infrastructure such as public transport, water supply, and waste management. Similarly, too little rainfall can disrupt water supply in urban areas. In rural areas of Selangor where crops such as palm and rubber are grown as well as livestock rearing, these extreme rainfall events can be debilitating. Too little or too much rainfall can cause crop failure. Literature reviewed showed mitigation measures such as changing types of crops or crop cycles were not adequate. These challenges make accurate rainfall prediction an essential strategy in planning and management within the Selangor state government. It is these challenges that were the main motivation of this study. This study aims to investigate if machine learning models can be used to produce accurate rain forecasts. These forecasts will be extremely useful to state government planners.

### Data Collection

A dataset consisting of five variables which are date, average temperature, wind speed, relative humidity and precipitation will be used. Use of these variables is well established in the literature. The target variable will be precipitation and the main objective of this study is to evaluate performance of machine learning models in predicting this variable. The predictors will be the other variables except date. The date variable will be useful in building time series models such as ARIMA. The selected dataset consists of daily observations covering the period between 2012 and 2020.

### Data Preprocessing

The selected dataset is expected to have some data quality issues. Exploratory data analysis will be used to identify missing values, values that are not within the expected range, and to understand the distribution of variables. Any missing values will be replaced with the mean value to avoid altering the distribution of variables. Any values that are not withing the expected range will be dropped in the analysis. To ensure all variables have an equal contribution to the model, each variable will be normalized. This will ensure all variables have a common range. In addition, the original daily data were combined into weekly data to reduce noise and show bigger trends in climate behaviour. A ratio of 80% to 20% will be used to split the dataset into train and test subsets. The train subset will be used for model training while the test subset will be used for model evaluation. These principles are well established in reviewed literature.

### Model Training

The models that will be investigated in this study are: artificial neural networks, support vector machines, de- cision trees, multiple linear regression, K-nearest neighbour, random forests, gradient boosting, and ARIMA. With the exception of linear regression all the other models have a set of parameters that will need to be tuned to achieve the highest prediction accuracy. These parameters are discussed for each model.

The artificial neural network has three architectural parameters that specify the general structure. They are layers, neurons in each layer, and activation functions. The layers and number of neurons will be used to achieve a balance between overfitting and long training time. Activation functions such as ReLu, Tanh, and Sigmoid will be used to capture non-linear patterns in the data. Various training parameters such as learning rate, batch size, epochs, and optimization methods such as SGD, RMSprop, and Adam will be examined to understand their influence on model accuracy. The dropout rate, L1, and L2 will be used to control overfitting.

The hyperparameters of a support vector machine that will need tuning are kernel, regularization, and epsilon. A non-linear relationship is expected in the data. Therefore, only radial basis and polynomial kernels will be examined. The regularization parameter will be tuned to control overfitting in the model. Epsilon will be tuned to control prediction accuracy.

The K-nearest neighbour hyperparameters that will be tuned are neighbours and distance metrics. The number of neighbours will be used to control overfitting. Various distance metrics such as Euclidean, Man- hattan, and Minkowski will be examined.

The random forest hyperparameters that will be tuned are: maximum depth, samples per leaf/tree, maximum features/leaf nodes, and split criterion. Tuning will ensure the model adequately captures the relationships in the data while avoiding overfitting or underfitting.

Gradient boosting parameters such as trees, learning rate, depth, split, subsampling, and features will be tuned to minimize overfitting and maximize prediction accuracy.

An ARIMA model requires optimal identification of p, d, and q parameters. Visual inspection and stationarity tests will be used to identify an optimal differencing order. The autocorrelation and partial autocorrelation functions will be used to identify optimal p and q parameters.

The R statistical software will be used for exploratory data analysis and model training. This software was selected because it is freely available and provides extensive data visualization and algorithm capabilities.

### Model Evaluation and Comparison

Three model evaluation metrics which are Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the Coefficient of Determination (*R*2) will be used to examine performance of models under investigation. RMSE captures the square root of the average squared differences between predicted and actual observa- tions. It shows the extent of large errors and is useful for identifying large deviations in rainfall predictions.

RMSE is easy to interpret as it is expressed in units of the response variable but has the limitation of not adequately capturing the influence of outliers. The formula for RMSE is: A square and square equation

AI-generated content may be incorrect.

A white paper with black text

AI-generated content may be incorrect.  
  
MAE captures the average difference in the absolute predicted and actual values. This provides a simple measure of prediction accuracy. MAE differs from RMSE as it considers all errors equal, making it robust against outliers. The formula for MAE is:

Where:

ˆ *yi*: Actual of observation i

ˆ *y*ˆ*i*: Prediction of observation i

ˆ *n*: Number of observations

ˆ Σ: Summation from 1 to i

A mathematical equation with numbers and symbols

AI-generated content may be incorrect.

*R*2 captures the extent to which the model explains the variation in the target variable. An *R*2 value close to 1 shows the model is very good at capturing a high degree of the variation, while a value close to zero is indicative of poor predictive performance. The formula for *R*2 is:

A mathematical equation with numbers and symbols

AI-generated content may be incorrect.

Where:

* yi: Actual of observation i
* yˆi: Prediction of observation i
* n: Number of observations
* Σ: Summation from 1 to i

These metrics will be very helpful in understanding the models under investigation. The MAE and RMSE will provide a quantitative value that indicates the difference between actual and predicted rainfall values. This will be useful in identifying the model that provides the best accuracy. *R*2 indicates the extent of model overfitting or underfitting. Therefore, comparison of these three metrics will provide a comprehensive performance evaluation.

Tables will be used to present the performance metrics of each model. This will facilitate easy comparison of the various models.

### Deployment

The selected machine learning model will be deployed as a prototype application to demonstrate its practical use. This application could be integrated into an early warning system or a web-based platform to provide real-time rainfall forecasts for stakeholders such as farmers, city planners, and disaster management author- ities. Deployment may involve creating a Python-based application with APIs to deliver actionable insights effectively.

# Expected Outcomes and Conclusions

## Introduction

This chapter will present the expected outcomes from the study. After carefully following the methodology developed earlier all study objectives will be achieved. The broad objective of the study is to investigate the potential of using machine learning in planning and management of extreme rainfall events in Selan- gor. Insights obtained from machine learning predictions will be used for agriculture, disaster, and water management planning.

## Expected Outcomes

This study is expected to meet its objectives. The first objective is to employ machine learning for rainfall prediction. This objective has been addressed through a comprehensive review of existing literature, which demonstrates the effectiveness of machine learning algorithms such as artificial neural networks, support vector machines, random forests, linear regression, K-nearest neighbours, gradient boosting, and ARIMA in forecasting rainfall. The literature also emphasizes the importance of practices such as data quality checks, data normalization, and appropriate train/test splits for ensuring model accuracy. Additionally, widely used evaluation metrics including RMSE, MAE, and R-squared will be adopted in this project to assess model performance.

The second objective will be to train identified machine learning algorithms using the data specified in the methodology. This objective has not been achieved. The methodology specified earlier will be followed in training each of the selected models. It is expected careful tuning of parameters will train models that balance computational cost, accuracy, and overfitting.

The third objective will be to identify the machine algorithm that provides the highest accuracy in rainfall prediction. This objective has not yet been met and it will only be achieved after examining results from objective 2. After training the models on the train subset, the performance of the models on the test subset will be examined using evaluation metrics and test subset. It is expected comparison of evaluation metrics will identify the algorithm with the highest accuracy.

## Conclusions

In conclusion, this research will build and evaluate machine learning models capable of accurately forecasting rainfall in Selangor. Using weather data and appropriate machine learning algorithms it is expected this study will identify a machine learning algorithm that can be incorporated into an early warning system. Such an early warning system will be critical to success of agriculture, infrastructure, and water management planning within Selangor. This study will demonstrate the value and limitations of using machine learning algorithms in rainfall prediction.

The findings are expected to provide actionable insights for various stakeholders, enabling better resource management, flood prevention, and agricultural planning. However, just like any other study this study will also have limitations. These limitations will only be fully clear after the project is completed. The findings of this study will then require interpretation in consideration of limitations.

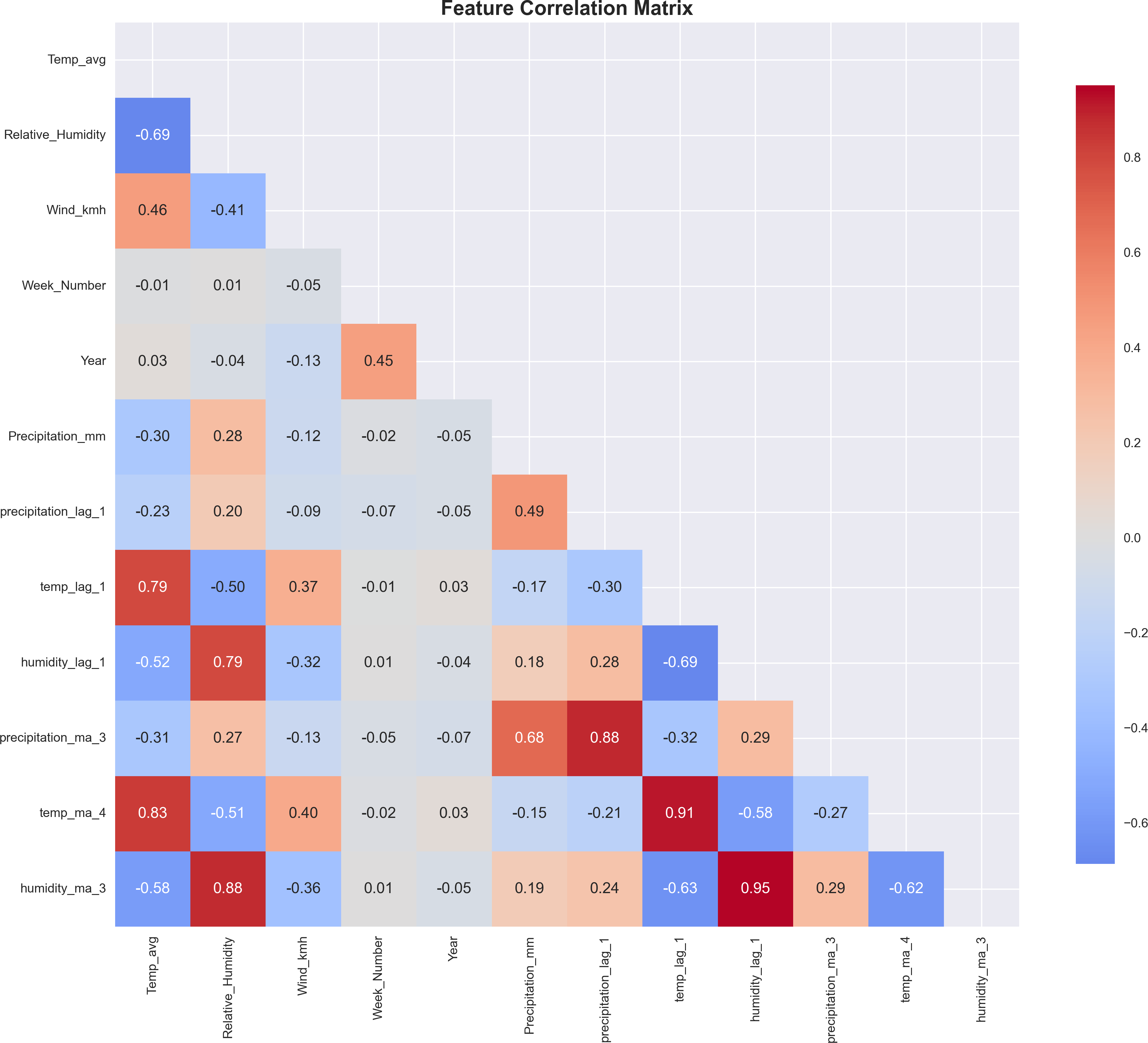


Figure 1: Correlation Matrix: The correlation matrix shows some important relationships between the weather variables. Temperature and humidity have a strong negative relationship, meaning when the tem- perature goes up, humidity usually goes down. Temperature also has a moderate positive link with wind speed, so higher temperatures often come with stronger winds. There is a weak negative connection between temperature and rainfall, suggesting that hotter days tend to have less rain. Rainfall and the “Rain Today” variable have a moderate positive link, which makes sense since more rain usually means it rained that day. The week and year don’t strongly affect the other variables, but they may still help track changes over time. Overall, temperature, humidity, and wind are useful for predicting rainfall.

# Acknowledgements

This research was supported by the Selangor State Government and Universiti Malaysia Pahang Al-Sultan Abdullah.

# References

1. Alam. Technological solutions for agricultural challenges. *Journal of Agricultural Technology*, X(Y):Z, 2021.
2. Z. B. Bouall`egue, M. C. A. Clare, L. Magnusson, E. Gasc´on, M. Maier-Gerber, M. Janouˇsek, M. Rodwell,

F. Pinault, J. S. Dramsch, S. T. K. Lang, B. Raoult, F. Rabier, M. Chevallier, I. Sandu, P. Dueben,

M. Chantry, and F. Pappenberger. The rise of data-driven weather forecasting: A first statistical as- sessment of machine learning–based weather forecasts in an operational-like context. *Bulletin of the American Meteorological Society*, 105(6):E864–E883, 2024.

1. Y. Kassem, H. G¨ok¸ceku¸s, H. C¸ amur, and E. Esenel. Application of artificial neural network, multiple linear regression, and response surface regression models in the estimation of monthly rainfall in northern cyprus. *Desalination and Water Treatment*, 215:328–346, 2021.
2. S. Kundu, S. K. Biswas, D. Tripathi, R. Karmakar, S. Majumdar, and S. Mandal. A review on rainfall forecasting using ensemble learning techniques. *e-Prime - Advances in Electrical Engineering Electronics and Energy*, 6:100296, 2023.
3. Nomadseason. Selangor climate, 2025. Accessed: January 18, 2025.
4. K. Ray, S. Balachandran, and S. K. Dash. Challenges of forecasting rainfall associated with tropical cyclones in india. *Meteorology and Atmospheric Physics*, 134(1), 2021.
5. S. a. A. Talib, W. M. R. Idris, L. J. Neng, T. Lihan, and M. Z. A. Rasid. Irregularity and time series trend analysis of rainfall in johor, malaysia. *Heliyon*, 10(9):e30324, 2024.
6. The National Oceanic and Atmospheric Administration. The challenges and complexities of weather forecasting, 2024.