PREDICTING FUTURE CHANGE OF SEASONAL RAINFALL AND TEMPERATURE IN BANGLADESH USING MACHINE LEARNING MODELS

BY

NUJHAT TABASSUM
ID: 201-15-3363

FINAL YEAR DESIGN PROJECT REPORT

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

MS. TANIA KHATUN
Assistant Professor
Department of Computer Science and Engineering
Daffodil International University

Co-Supervised By

MS. SHARUN AKTER KHUSHBU
Lecturer (Sr. Scale)
Department of Computer Science and Engineering
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

July 2024

APPROVAL

This Project titled **"Predicting Future Change of Seasonal Rainfall and Temperature in Bangladesh Using Machine Learning Models"**, submitted by **Nujhat Tabassum** to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on **13-07-2024.**

BOARD OF EXAMINERS

Board Chairman

Dr. S.M Aminul Haque (SMAH)
Professor & Associate Head
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner 1

Ms. Nazmun Nessa Moon (NNM)
Associate Professor
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

13.7.24

Internal Examiner 2

Mr. Deawan Rakin Ahamed Remal (DRAR)
Lecturer
Department of CSE
Faculty of Science & Information Technology
Daffodil International University

13.07.2024

External Examiner

Dr. Abu Sayed Md. Mostafizur Rahman (ASMR)
Professor
Department of CSE
Jahangirnagar University

DECLARATION

We hereby declare that this project has been done by us under the supervision of **Ms. Tania Khatun, Assistant Professor, Department of Computer Science and Engineering, Daffodil International University**. We also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Tania Khatun 11-07-24

Name: Ms. Tania Khatun
Designation: Assistant Professor
Department of CSE
Daffodil International University

Co-Supervised by:

Name: Sharun Akter Khushbu
Designation: Lecturer (Senior Scale)
Department of CSE
Daffodil International University

Submitted by:

Nujhat Tabassum 01-08-24

Name: Nujhat Tabassum
ID: 201-15-3363
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty for His divine blessing making it possible for us to complete the final year project/internship successfully.

We are grateful and wish our profound indebtedness to **Ms. Tania Khatun**, **Assistant Professor**, **Department of CSE, Daffodil International University, Dhaka**. Deep Knowledge & keen interest of our supervisor in the field of Computer Science & Engineering to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts, and correcting them at all stages have made it possible to complete this project.

We would like to express our heartiest gratitude to the **Head of the Department of CSE Dr. Sheak Rashed Haider Noori**, for his kind help in finishing our project and also to other faculty members and the staff of the Department of CSE, Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discussion while completing the course work.

Finally, we must acknowledge with due respect the constant support and patience of our parents.

ABSTRACT

Understanding the dynamic shifts in rainfall and temperature holds immense significance for countries like Bangladesh grappling with critical climate conditions. Natural disasters such as floods, droughts, and cyclones are intricately tied to variations in these climatic factors. Therefore, the ability to predict future patterns of rainfall and temperature is crucial for effective disaster management. While numerous methods exist for forecasting these conditions, the exploration of predicting future changes in seasonal rainfall and temperature using diverse machine learning models and assessing their accuracy remains unexplored. This thesis aims to fill this gap by systematically comparing six machine learning models such as Linear Regression, Polynomial Regression, Decision Tree, Random Forest, Adaboost and Gradient Boosting Machines to forecast the seasonal variations in rainfall and temperature for Bangladesh. The emphasis is on identifying the most accurate model through rigorous comparisons. The selected optimal machine learning model is subsequently employed to project seasonal rainfall and temperature conditions for the year 2025 to 2074. The analysis extends further to examine the percentage change in seasonal rainfall and change in average temperature between the years 2024 and 2050. By adopting this approach, the research contributes to advancing the understanding of climate prediction methodologies, particularly in the context of a country facing acute climate challenges like Bangladesh.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
Introduction

1.1 Overview

Changes in seasonal temperature and rainfall patterns present considerable challenges to ecosystems, societies, and economies globally. Accurately predicting these changes is crucial for effective decision-making, particularly in vulnerable regions such as Bangladesh. Understanding seasonal climate variations helps in planning and mitigation strategies for agriculture, water resources, public health, and infrastructure. This study aims to forecast future changes in seasonal rainfall and temperature using advanced machine learning models. Machine learning techniques offer the potential for more accurate predictions by analyzing large datasets and identifying complex patterns. By leveraging historical climate data, the study seeks to develop predictive models that can provide reliable forecasts. These predictions can inform policymakers and stakeholders, enabling them to implement proactive measures. The research addresses the need for improved climate predictions to enhance resilience against adverse weather impacts. Ultimately, the study contributes to the broader effort of mitigating the effects of climate change on vulnerable populations. The findings underscore the importance of integrating technological advancements into climate science for better preparedness and adaptation strategies.

1.2 Background and Present State

Climate change poses significant challenges for countries worldwide, and Bangladesh is particularly vulnerable due to its geographic location and socio-economic conditions. The dynamic shifts in rainfall and temperature in this region are critical factors contributing to natural disasters such as floods, droughts, and cyclones, which have profound implications for agriculture, water resources, and disaster management. Therefore, predicting future patterns of seasonal rainfall and temperature is paramount for effective climate adaptation and mitigation strategies. While traditional methods have been employed for such predictions, there is a growing interest in utilizing advanced machine learning models to enhance the accuracy and reliability of these forecasts.

Recent studies have demonstrated the potential of machine learning techniques in predicting climatic variables with greater precision. For instance, Sashank et al. [1] explored the application of artificial neural networks (ANNs) for rainfall prediction, highlighting the complex relationships between climate variables. Similarly, Rudrappa et al. [2] provided a comprehensive comparison of various machine learning algorithms, such as Decision Tree, Support Vector Machine (SVM), K-nearest Neighbor (KNN), and Multi-Layer Perception (MLP), in the context of rainfall forecasting. These studies underscore the importance of leveraging machine learning models to improve the predictability of climatic events, which can significantly benefit sectors like agriculture and water resource management.

Despite these advancements, there remains a gap in the literature regarding the long-term prediction of seasonal rainfall and temperature changes specifically for Bangladesh. The unique climatic challenges faced by Bangladesh necessitate tailored forecasting models that can accurately project future scenarios. Previous research, such as the works of Mekanik et al. [3] and Oswal et al. [4], have primarily focused on short-term predictions and specific regions, which may not fully capture the seasonal variations pertinent to Bangladesh. This thesis aims to address this gap by systematically comparing different machine learning models to forecast seasonal rainfall and temperature for Bangladesh from 2025 to 2074, thereby providing valuable insights for climate resilience planning.

The present state of the thesis builds on an extensive background study that highlights the critical need for accurate climate predictions, especially for regions like Bangladesh, which are highly vulnerable to the impacts of climate change. Climate change has exacerbated the frequency and severity of natural disasters such as floods, droughts, and cyclones, making it imperative to understand and predict changes in seasonal rainfall and temperature. Traditional prediction methods have provided valuable insights but often fall short in accuracy and reliability due to the complex and dynamic nature of climatic variables.

The thesis will build upon the methodologies discussed in existing literature, incorporating advanced models and optimization techniques to enhance prediction accuracy. For instance, Azad et al. [5] suggested combining ANFIS with optimization algorithms to improve temperature forecasting, a concept that can be extended to rainfall prediction as well. Additionally, Kumar et al. [6] and Oyana [7] highlighted the effectiveness of Random Forest and k-Nearest Neighbor algorithms in climate predictions, which will be critically assessed in this study for their applicability to Bangladesh's climatic context. By integrating these models, the research will aim to identify the most robust and accurate forecasting approach.

Ultimately, the selected optimal machine learning model will be employed to project seasonal rainfall and temperature conditions for the specified period. The selected machine learning models will be used to predict the future seasonal rainfall and average temperature for the year 2025 to 2074. This projection will facilitate an in-depth analysis of the percentage change in seasonal rainfall and change in average temperature in degree Celsius between 2024 and 2050, providing a clearer picture of the potential climatic shifts in Bangladesh. By adopting a comprehensive and methodologically rigorous approach, this research aims to contribute to the broader understanding of climate prediction methodologies, offering valuable insights for policymakers, researchers, and stakeholders involved in climate adaptation and disaster management in Bangladesh.

This study will utilize six machine learning models: Linear Regression, Polynomial Regression, Decision Tree, Random Forest, Adaboost and Gradient Boosting Machines. The primary objective is to identify the most accurate model for predicting future climatic conditions. Each model will be rigorously tested and evaluated based on its accuracy in forecasting seasonal rainfall and average temperature. The model demonstrating the highest accuracy will be selected for further analysis. Upon selection, this optimal model will be employed to predict future seasonal rainfall and average temperature for Bangladesh. This approach aims to provide precise and reliable forecasts, essential for effective disaster management and climate adaptation strategies. The use of diverse models ensures a comprehensive analysis and comparison, enhancing the robustness of the predictions. By focusing on accuracy, the study aims to deliver the most reliable insights into future climatic trends. These predictions will cover the period from 2025 to 2074, offering valuable data for long-term planning. The findings are expected to significantly contribute to the understanding and management of climatic challenges in Bangladesh.

Linear Regression Model

Linear regression is a method in supervised machine learning that seeks to establish a straight-line relationship between a target variable and one or more predictor variables. It aims to model and quantify the association between these variables in a way that allows for predictions and insights into their mutual dependencies.

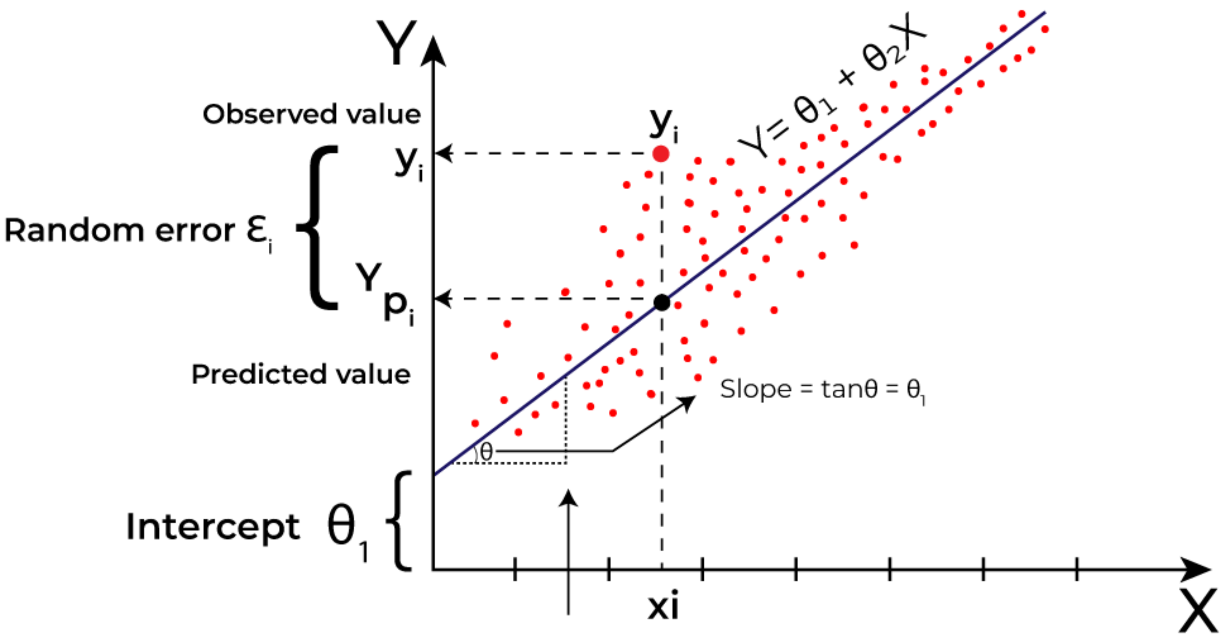

Figure 1.1: Linear Regression Model

Polynomial Regression Model

Polynomial Regression is an extension of linear regression that introduces a more flexible approach to modeling the connection between an independent variable (x) and a dependent variable (y). Instead of assuming a straight-line relationship, polynomial regression accommodates curved patterns by employing an nth-degree polynomial equation. This method allows for a better representation of non-linear associations between x and the expected value of y given a specific x value.

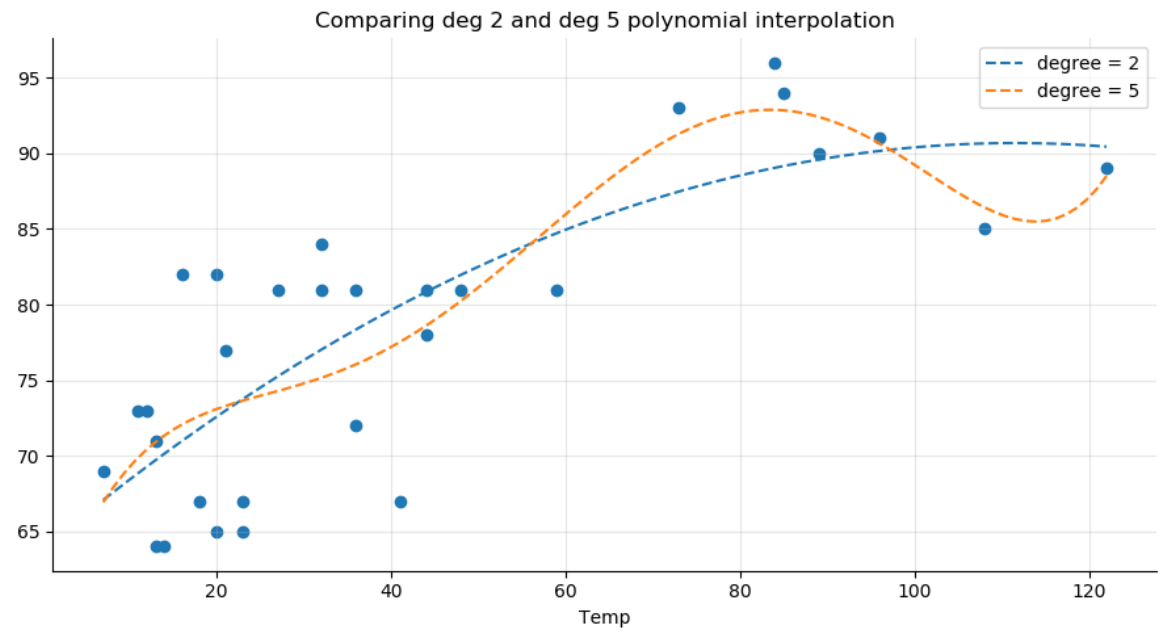

Figure 1.2: Polynomial Regression Model

Decision Tree Model

A decision tree is one of the most powerful tools of supervised learning algorithms used for both classification and regression tasks. It builds a flowchart-like tree structure where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label.

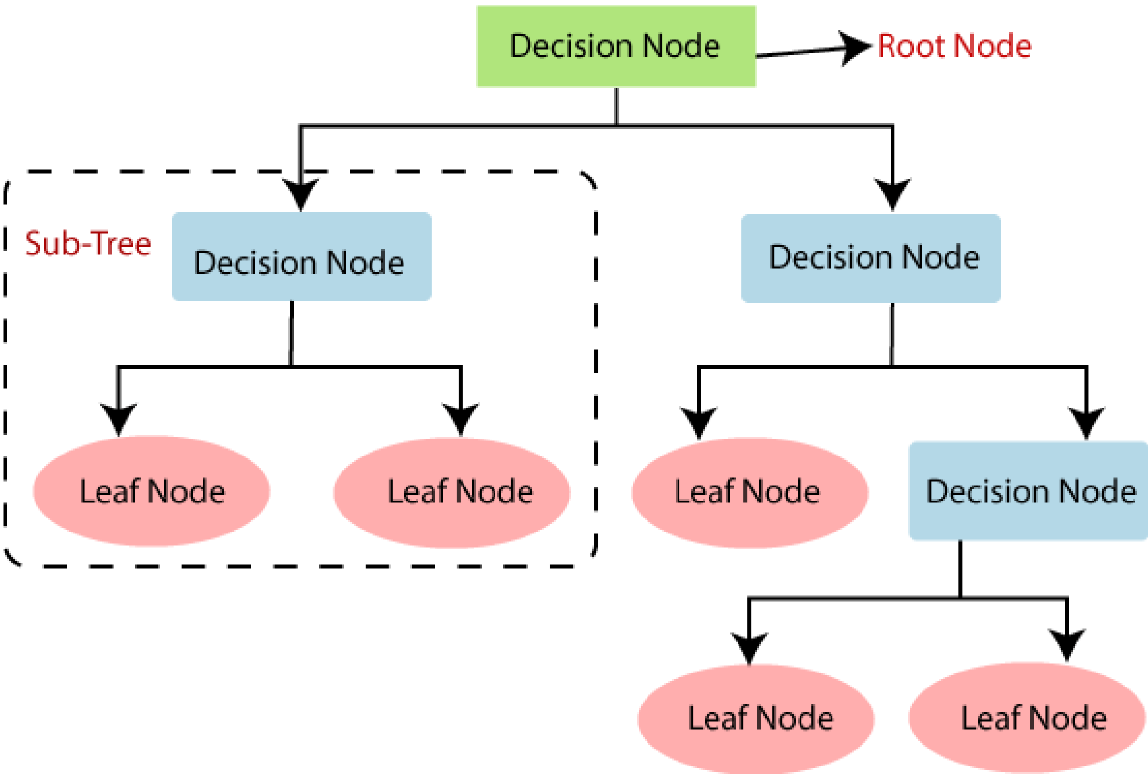

Figure 1.3: Decision Tree Model

Random Forest Model

Random Forest serves as a guiding force in the realm of supervised machine learning, adept at tackling both classification and regression challenges. It operates by constructing an ensemble of decision trees, each drawn from a distinct, randomly chosen subset of the training data. As these diverse decision trees contribute their individual insights, the Random Forest method harmonizes their collective wisdom through a voting process. The amalgamation of these varied perspectives ultimately.

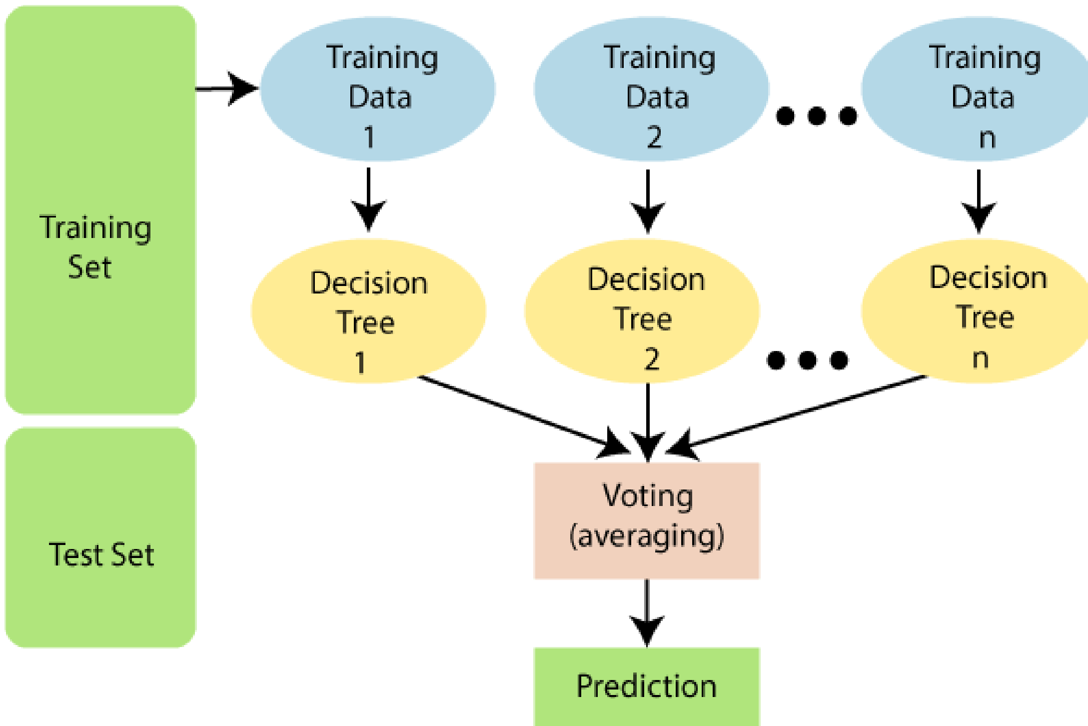

Figure 1.4: Random Forest Model

Adaboost

AdaBoost, short for Adaptive Boosting, is a powerful ensemble learning technique that combines multiple weak learners, typically decision trees, to form a strong predictive model. Developed by Freund and Schapire in 1996, AdaBoost works by sequentially fitting these weak learners on the training data, with each subsequent model focusing more on the instances that were previously misclassified. This is achieved by adjusting the weights of the training samples, giving higher weight to those that were harder to classify. The final model is a weighted sum of the weak learners, leading to a robust classifier that often performs better than individual models. AdaBoost is particularly effective in reducing bias and variance, making it a popular choice for classification tasks in various domains.

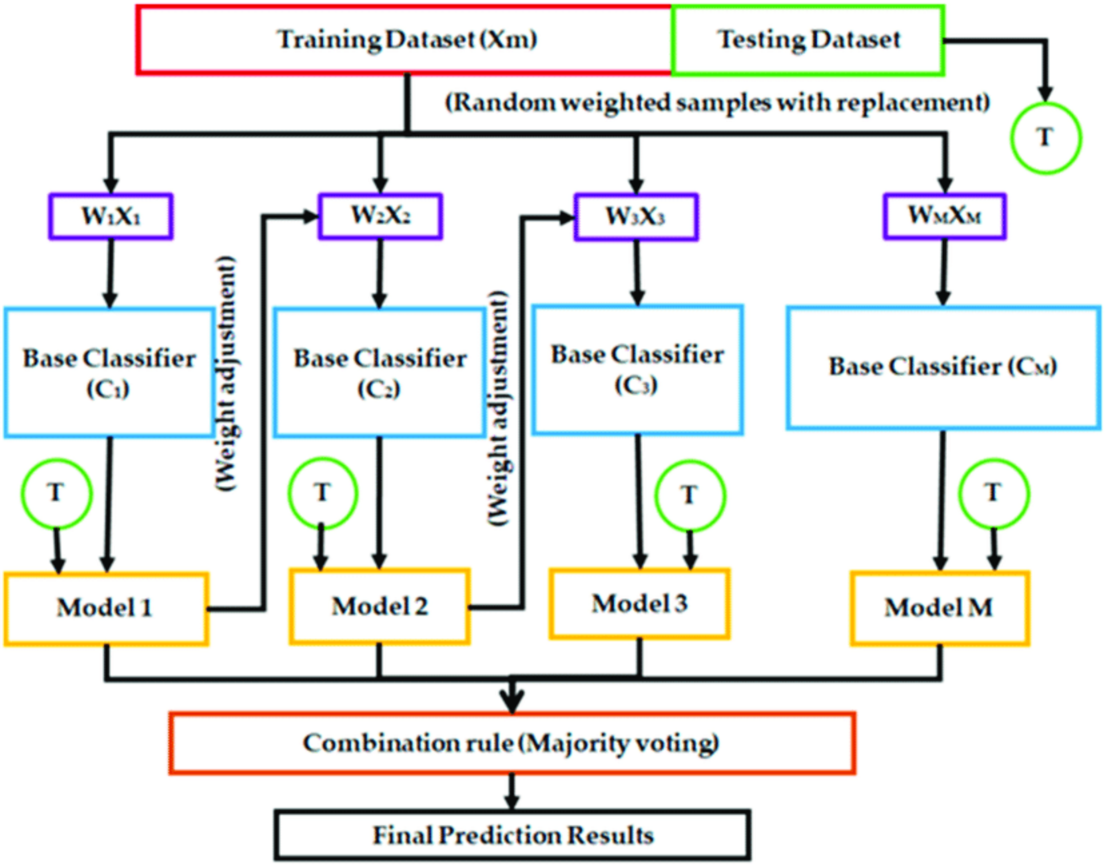

Figure 1.5: Adaboost Machine Learning Model

Gradient Boosting Machines

Gradient Boosting Machines (GBM) is a powerful ensemble learning technique that builds a robust predictive model by sequentially adding weak learners, typically decision trees, in a stage-wise fashion. Each new tree is trained to correct the residual errors of the previous trees, and this process is guided by gradient descent optimization to minimize a specified loss function. GBM is highly flexible and can be used for both regression and classification tasks. It excels in capturing complex patterns in the data by iteratively improving model accuracy, while also incorporating regularization techniques to prevent overfitting. Due to its high performance and ability to handle various data types, GBM is widely used in many practical applications, including finance, healthcare, and marketing.

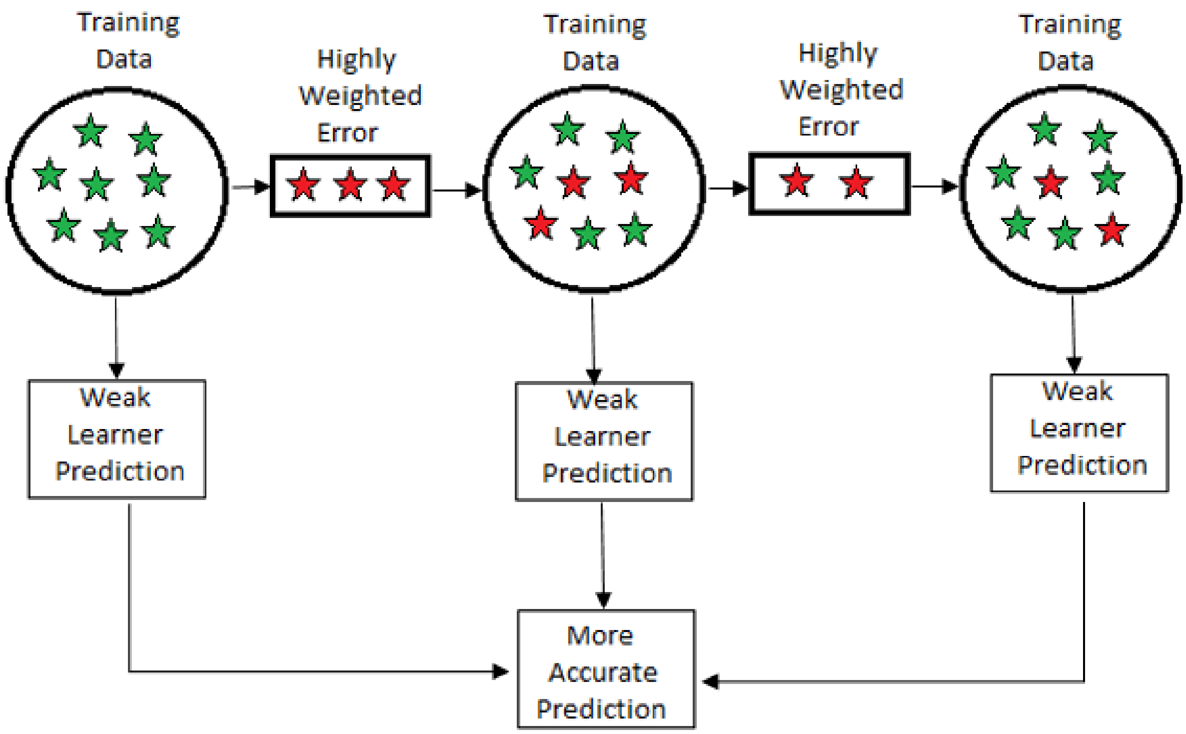

Figure 1.6: Gradient Boosting Machines

1.3 Problem Statement

The fundamental problem addressed in this thesis centers on the need to develop accurate and reliable models for forecasting future fluctuations in seasonal climate variables, particularly rainfall and temperature. As climate change intensifies, the capacity to predict these changes becomes crucial for devising adaptive strategies and policies. Reliable forecasts can aid in mitigating adverse impacts on agriculture, water resources, and infrastructure. This study employs advanced machine learning models to enhance prediction accuracy. By leveraging historical climate data, the research aims to create robust predictive tools. Ultimately, these models will support informed decision-making and resilience planning in the face of evolving climatic conditions.

1.4 Objectives

The primary objectives of this thesis are to develop and evaluate machine learning models for accurately forecasting future changes in seasonal rainfall and temperature in Bangladesh. By utilizing historical climate data, the research aims to identify patterns and trends that can enhance predictive accuracy. The study will compare the performance of various models, including linear regression, polynomial regression, decision trees, random forest, adaboost and gradient boosting machines, to determine the most effective approach. Additionally, the thesis seeks to address current gaps in the literature by standardizing methodologies and providing comprehensive comparative analyses. Through rigorous testing and validation, the research will assess model limitations and generalizability across different climatic regions within Bangladesh. The ultimate goal is to create reliable predictive tools that support adaptive strategies and policy development. This work intends to contribute to improved resilience against climate variability and its impacts on critical sectors such as agriculture, water resources, and infrastructure.

1.5 Scope and Limitations

The scope of this thesis encompasses the development and application of machine learning models to predict future changes in seasonal rainfall and temperature in Bangladesh. By leveraging historical climate data from 1951 to 2014, the study aims to create accurate and reliable predictive models using techniques such as linear regression, polynomial regression, decision trees, random forest, adaboost and gradient boosting machines. The research focuses on synthesizing insights from these models to identify trends and patterns in climate variables, which are crucial for developing adaptive strategies in vulnerable regions. Additionally, the study seeks to standardize methodologies for model evaluation and comparison, thereby enhancing the reliability and reproducibility of findings. This comprehensive approach aims to provide valuable tools for policymakers and stakeholders to mitigate the impacts of climate change on agriculture, water resources, public health, and infrastructure.

However, the thesis also faces certain limitations. One significant constraint is the quality and granularity of historical climate data, which may affect the accuracy of the predictive models. The study is also limited to the climatic conditions of Bangladesh, potentially restricting the generalizability of the findings to other regions with different environmental characteristics. Additionally, the complexity of climate systems means that models may struggle to account for all influencing factors, leading to potential inaccuracies. The thesis also acknowledges the challenges in ensuring computational efficiency and managing the large datasets required for training the machine learning models. Furthermore, the reliance on historical data may not fully capture future climate anomalies driven by unprecedented global changes. Despite these limitations, the research aims to provide a robust framework for climate prediction that can be refined and expanded in future studies.

1.6 Report Organization

The proposed report is organized with a detailed structure designed to lead readers through the research process, findings, and implications systematically. Each chapter fulfills a distinct role, enhancing the document's overall coherence and depth.

Chapter 1: Introduction

The introductory part of the thesis provides a comprehensive overview of the research's background, emphasizing the critical importance of predicting seasonal rainfall and temperature changes in Bangladesh due to its vulnerability to climate-induced disasters. It outlines the study's primary objectives and research questions, setting the stage for the detailed exploration of advanced machine learning models. The introduction also reviews relevant literature, situating the research within the broader context of climate science and highlighting gaps that this study aims to address. It briefly describes the methodologies employed, including the selection and evaluation of various machine learning models. Finally, the chapter underscores the significance of the study's findings for policymakers and stakeholders in climate adaptation and disaster management in Bangladesh.

Chapter 2: Background

The background chapter delves into a thorough examination of relevant literature and prior research. This section provides a comprehensive understanding of the theoretical foundations, methodologies, and technological advancements pertinent to predicting rainfall and temperature. It also covers the six machine learning models used and the concept of transfer learning. By establishing the context for the current study, this chapter underscores the existing knowledge gaps and sets the stage for addressing them through this research.

Chapter 3: Research Methodology

The research methodology chapter details the approach employed in conducting the study. It offers an overview of the research design, data collection techniques, model structures, and the reasoning behind the selected methodologies. Additionally, this chapter addresses ethical considerations and acknowledges any limitations that could influence the study's outcomes.

Chapter 4: Experimental Result

This chapter presents the model results obtained from the six machine learning models such as linear regression, polynomial regression, decision tree, random forest, adaboost and gradient boosting machines. Predicted values of seasonal rainfall and average temperature using the models selected based on the model accuracy are shown in this chapter. Moreover, this chapter contains analysis of change of rainfall in percentage and average temperature in degree Celsius with respect to current situation.

Chapter 5: Impact on Society

This chapter discusses the significant societal implications of using machine learning models to predict future changes in seasonal rainfall and temperature in Bangladesh. It highlights how accurate climate forecasts can aid policymakers in devising strategies to mitigate adverse impacts on agriculture, water resources, public health, and infrastructure. The chapter emphasizes the benefits for the agricultural sector, where precise climate predictions help farmers make informed decisions, thus enhancing food security and economic stability. It also covers the improvement in water resource management and public health preparedness by anticipating climate-induced changes. Overall, the research promotes climate awareness and encourages sustainable practices, contributing to the socio-economic stability and resilience of Bangladesh.

Chapter 6: Summary, Conclusion, Recommendation, and Implications for Future Research

This final chapter encapsulates the study's objectives, key findings, and their implications. It discusses the development and evaluation of machine learning models to predict seasonal rainfall and temperature changes in Bangladesh, emphasizing the anticipated climate shifts and their impact on agriculture, water management, and public health. The conclusions highlight the necessity for adaptive strategies and ethical communication of findings to stakeholders. Recommendations include adopting advanced machine learning models, incorporating additional climate variables, and refining predictions through regional climate models. The chapter underscores the importance of ongoing research and stakeholder collaboration to enhance climate resilience and inform future policy-making.

This structured layout ensures a logical progression of information, taking the reader step-by-step through the research journey, from the introduction to the broader implications and recommendations. It facilitates a comprehensive grasp of the research process and its potential impact on both academic and practical fronts. By employing machine learning models like linear regression, polynomial regression, decision tree, random forest, adaboost, gradient boosting machines, the study provides valuable insights into predicting future seasonal rainfall and average temperature. This organized approach helps in understanding the methodology and the significance of the findings, highlighting the practical applications and broader consequences of the research.

1.7 Summary

This study focuses on predicting future changes in seasonal rainfall and temperature using advanced machine learning models, addressing significant challenges posed by climate change. Accurate predictions are crucial for effective decision-making, especially in vulnerable regions like Bangladesh, impacting agriculture, water resources, public health, and infrastructure. By analyzing historical climate data from 1951 to 2014, the research aims to develop reliable predictive models, including linear regression, polynomial regression, decision trees, random forest, adaboost and gradient boosting machines. These models will help identify patterns and trends, providing valuable tools for policymakers and stakeholders to mitigate climate change impacts. Despite limitations such as data quality and the complexity of climate systems, the study strives to enhance predictive accuracy and contribute to climate resilience planning. Ultimately, this research aims to inform adaptive strategies and policy development, promoting better preparedness and adaptation to evolving climatic conditions.

CHAPTER 2
Literature Review

2.1 Overview

To ensure clarity and comprehension of the methodologies and core concepts explored in this research on "Predicting Future Changes in Seasonal Rainfall and Temperature in Bangladesh Using Machine Learning Models," it is crucial to establish a solid foundation by first defining key terms and outlining fundamental concepts. This foundation involves analyzing climate variability, which refers to natural fluctuations in weather patterns over time, using historical climate data collected from 1951 to 2014 across 188 locations. The study utilizes various machine learning models including linear regression, polynomial regression, decision tree, random forest, adaboost and gradient boosting machines to forecast seasonal rainfall and temperature, categorizing data into Premonsoon, Monsoon, Postmonsoon, and Winter seasons. Transfer learning techniques enhance model performance by leveraging knowledge from related domains. Adaptive agricultural practices and water resource management strategies are essential to mitigate the impacts of climate change, while public health initiatives address health risks associated with shifting climate conditions. Ethical considerations focus on responsible data handling and transparent communication of findings, while sustainability principles guide efforts towards long-term resilience and resource management. These terminologies underpin the study's methodology and implications for enhancing climate resilience in Bangladesh.

2.2 Related Works

Sashank, C. et al., [1] The paper concentrates on the establishment of a rainfall predicting tool using neural networks trained to understand complicated relationships between climate variables and rainfall on big datasets of meteorological data. The study does not lucidly state the limitations of the suggested rainfall predicting tool and the neural network models used for predicting. It offers a tool which uses neural networks to precisely forecast rainfall grounded on historical meteorological data, allowing farmers to schedule their harvests, water resources engineers and executives to distribute resources, and general public in the corresponding area to become prepared and take necessary steps for potential floods, droughts and other natural disasters. In this paper, Artificial Neural Networks (ANN) models have been used for Rainfall prediction that can be further examined considering the parallel processing competencies of ANNs and the huge number of interconnected neurons.

Gujnatti Rudrappa et al., [2] The paper highlights the importance for weather impact researches and agriculture as superfluous or inadequate rainfall can be destructive to yields. Machine learning (ML) models can be used to forecast rainfall and its influence on crop. This study does not deliver a detailed dissertation on the certain limits of the proposed machine learning models for predicting the rainfall. This paper also offers a thorough comparison among various machine learning algorithms for rainfall prediction. This comparison allows scientists and researchers to comprehend the strong points and limitations of the models. This paper offers more complex research on the comparison of miscellaneous machine learning algorithms for rainfall prediction like Decision Tree algorithms, Support Vector Machine (SVM), K-nearest Neighbor (KNN) and Multi-Layer Perception (MLP).

Mekanik, F. et al., [3] This paper gives a prediction-based idea where regression and artificial neural network are used to predict the rainfall. This idea concentrates on only the rainfall but it is not aiming on temperature nor greenhouse gases.

Oswal, N. et al., [4] The paper represents comparative research on using machine learning models to forecast rainfall in principal municipalities of Australia grounded on meteorological data, concentrating on modeling inputs, methods, and pre-processing techniques. The research concentrates on principal municipalities in Australia, which may not be representative of rainfall patterns of other localities and countries. The study aims to the difficult and challenging task of the forecasting of rainfall having considerable influences on social life. Findings from the study demonstrated that Gradient Boosting with a learning rate of 0.25 performed best in terms of accuracy, while Random Forest and Decision Tree performed worst in terms of coverage in Experiment 1 with the original dataset.

Azad, A. et al., [5] This study suggested an innovative procedure to forecast air temperature through combining the ANFIS and optimization algorithms. The suggested procedure presented an increase in the accuracy of the forecasting.

Kumar, V. et al. [6] This paper recognizes the challenge of precise long-term rainfall prediction through presenting the limitations of long-term rainfall prediction. This paper concentrates on the significance of rainfall prediction in miscellaneous areas such as power generation, agriculture and research, and goals to establish a rainfall prediction system using machine learning algorithms. Random Forest presents the best model accuracy with 88.21 score while Decision Tree shows the worst model accuracy with 73.67 score. K-nearest Neighbor (KNN) (n=27) presents the accuracy with 87.36 score. It offers solutions to the challenging issues of rainfall prediction in the field of power generation, agriculture and research. Further study could be conducted to consider the use of other machine learning algorithms like and predicting the amount of precipitation and comparing its performance with random forest and algorithms and K-nearest neighbor machine learning model.

Oyana, T. J., [7] This paper represents the k-Nearest Neighbor (kNN) regression operating on the concept of spatial dependency. The scheme means the value of the nearest k neighbors around an event to forecast its value through rewarding closer events and penalizing far events. The peak number of neighbors (k) is established in training the method since it dominates the predictive values and computational effectiveness.

Liew, C.M. et al., [8] This paper centers on daily rainfall prediction in Ethiopia that is essential for water sectors and agricultural production through using machine learning models. This study does not offer information about certain limits and challenges met during the implementation of machine learning algorithms for rainfall prediction. The paper indicates that further work may contain the use of big data analysis for rainfall prediction, encompassing sensor and climate data.

Shri Sandeep et al., [9] This paper represents rainfall prediction as a difficult and states that precise predictions can assist in taking necessary steps as safety measures in water management strategies in different water related fields. The study does not deliver certain features about the machine learning models used for rainfall prediction, making it challenging to evaluate the accuracy of the models. It keeps a delegated survey and relative analysis of various neural networks for rainfall prediction which underlines the merits of RNN, FFNN and TDNN compared to other methods. Random Forest model presents the best accuracy with 87.76 score while Artificial Neural Networks Model shows the worst accuracy with 84.80 score. The accuracy scores of the other two methods such as Linear

Regression model and Naïve Bayes model represent 87.15 and 85.01 respectively.

Singla, P. et al., [10] This paper concentrates on the significance of rainfall prediction in agriculture and emphasizes the use of machine learning models such as ARIMA, Logistic Regression (LR), Artificial Neural Network (ANN), and Support Vector Machine (SVM) in finding the potential of these techniques in agriculture. This study does not review the challenges in executing machine learning models for rainfall prediction. Random Forest (RF) and XGBOUT models both show the prediction accuracy 0.95 and Decision Tree Model presents the accuracy score 0.85.

Alizadeh, M. J. et al., [11] This paper represents one of the important concepts for forecasting the rainfall. In this study, Artificial Neural Network and Support Vector Regression (SVR) models are used. This concept only concentrates on rainfall but it does not offer any idea about temperature.

Byun, J. et al., [12] This paper offers a new method for rainfall prediction using cloud image data collected from IoT sensors by using convolutional neural networks (CNNs) and binary classification models. Cloud image data collected from IoT sensors at specific locations in Seoul, Republic of Korea was used to conduct the study that limits the generalizability of the outcomes to other regions or climates. The paper represents a new research approach for predicting rainfall using cloud image data that can deliver valuable understandings about microscale metrological information. The recommended CNN-based image-value model forecasts rainfall with an average mean square error (MSE) of 85.59%.

Poornima, S. et al., [13] This paper represents an intensified LSTM based RNN model for rainfall prediction where training and testing are conducted using a standard rainfall dataset. This study does not deliver a thorough discussion on the limits of the suggested intensified LSTM model for rainfall prediction. This paper suggests an intensified LSTM based recurrent neural network for rainfall prediction that can help in the policy-making practice of groups responsible for the prevention of disasters. Performing prescriptive analytics is resulted through analysis of positive and negative impacts of rainfall prediction.

Rahman, A. et al., [14] The paper offers a real-time rainfall forecasting scheme for smart cities using a machine learning fusion technique that outperforms other models and algorithms. The study does not consider the certain difficulties met during the selection of the six supervised machine learning techniques used in the recommended outline. The suggested structure applies a fusion technique which merges the forecasting accuracies of different machine learning techniques, resulting in advanced rainfall forecasting accuracy. The outline of the thesis represented in the study will be prolonged in the future by investigating the fusion of ensemble machine learning techniques on more distinct datasets. In this study, Support Vector Machine (SVM) shows the best model accuracy with 93 percent score while K-nearest Neighbors (KNN) presents the 90 percent model accuracy which is the worst model efficiency among the used models. The other two models such as Decision Tree and Naïve Bayes show 91 percent and 92 percent model accuracy respectively.

Barrera-Animas et al., [15] This paper implies the significance of Rainfall prediction for flood prediction and monitoring contaminant concentration levels and explores the use of Machine Learning algorithms for effective and budget-wise rainfall prediction applications. The study does not discuss the challenges of the models used for rainfall prediction like the possibility for overfitting and the generalizability of the models to distinct areas or time frames. Relative analysis of rainfall estimation models using Machine Learning models and Deep Learning designs and architectures delivers understandings of their performance and appropriateness for rainfall prediction applications. Stacked-LSTM achieved Loss values between 0.0014–0.0001, RMSE values in the range of 0.0375–0.0084, MAE values between 0.0071–0.0013, and RMSLE values ranging between 0.0157–0.0037. On the other hand, Bidirectional-LSTM obtained values ranging between 0.0014–0.0001, 0.0377– 0.0099, 0.0072–0.0015, and 0.0111–0.0044 for Loss, RMSE, MAE, and RMSLE, respectively.

Dutta, K. et al., [16] This paper uses machine learning and neural networks for comparing two approaches to predict and identify more precise rainfall estimation. This study does not keep descriptions of the particular neural network and machine learning models used for rainfall prediction, confining the reproducibility and validation of the results and findings. The paper conducts a comparative study of different methods for rainfall prediction, with a particular focus on machine learning models and neural networks, and offer understandings of their accuracy and efficiency. The error value of MSE is 0.08663 and the error value of MAE is 0.249793. This paper claims that further study should focus on forming models and algorithms which tests daily data instead of cumulative data sets to improve the accuracy of rainfall predictions.

Salcedo-Sanz, S. et al., [17] This paper offers a prediction-based study. In this article, only prediction of temperature is center but it does not give any idea about the prediction of rainfall. The physical factors are not concentrated which are accountable for the Global Warming. The study uses the Support Vector Regression (SVR) and Multilayer-Perceptron methods.

Umay Shah et al., [18] This paper concentrates on using statistical methods and machine learning models to forecast and estimate climate parameters where the random forest model presents the best classification accuracy for forecasting rainfall for the later season. This study keeps perceptions of meteorological parameters like rainfall, temperature and humidity which are significant for various industries as well as scientists and researchers.

Joseph et al., [19] This paper goals to investigate the application of data mining techniques for rainfall prediction, specially clustering and classification. The study uses clustering and classification techniques for rainfall prediction that can provide constructive understandings and patterns of the dataset. The paper presents a model based on k-means clustering technique with supervised data classification technique for rainfall prediction using wide-ranging meteorological variables in a river basin.

Kumar, P. et al., [20] This paper shows that temperature has an obvious effect on rainfall which is responsible for generating drought event. This study focuses on predicting drought event. To forecast drought event, different machine leaning models and algorithms are used in this research. This research claims that those machine learning approaches would assist in predicting drought event.

Wang, Y. et al., [21] This study utilized two distinct machine-learning models named as k-Nearest Neighbors (kNN) and a Pattern Approximate Matching algorithm for facilitating the forecasting of short-term air temperature. The k-Nearest Neighbors model is a non-parametric and instance-based procedure which categorizes a data point grounded on the majority class of its k-nearest neighbors in the feature space. In contrast, the Pattern Approximate Matching algorithm encompasses recognizing and matching patterns within the dataset to make forecasting, suggesting an exceptional method to capture difficult relationships in the data for precise short-term air temperature prediction. Through blending these two methodologies, the study intended to increase the accuracy and efficiency of short-term air temperature forecasting.

Bindu, B. H., [22] This paper describes that Rainfall prediction is an influential aspect of climate prediction with inferences for water related fields, agriculture, flood mitigation, and disaster alertness. Machine learning models have evidenced to be a strong tool for improving the precision of rainfall predictions. Various models are used in this rainfall forecast such as Random Forest (RF), Decision Tree (DT), K-Nearest Neighbor (KNN), Multilayer Probe (MLP) and Extreme Gradient Boosting (XGB). Convenient information can be gathered from rainfall to plan long-term water management strategies and design suitable irrigation techniques.

Vishwakarma, H., [23] This paper assesses the performance of various Machine Learning algorithm such as Linear Regression (LR), Support Vector Regression (SVR), lasso and ElasticNet to predict annual global warming. This paper shows that CO2 is the most significant contributor causing temperature change, followed by CH4, N20 and SF6. After observing the analyzed and forecasted data of the greenhouse gases and temperature, the global warming can be degraded relatively within few years.

Azari, B., [24] This study took a local temperature time series as a dependent variable and a collection of related meteorology time series as independent variables and employed principal Machine Learning methods to them. Six machine learning models such as Linear Regression (LR), k-Nearest Neighbor (kNN), Support Vector Machine (SVM) and Artificial Neural Network (ANN), Random Forest (RF) and Adaptive Boosting (AdB) have been applied in this study. Genuine data was used to train each of the mentioned ML approaches to forecast the future temperature in the study area. The analysis disclosed that Artificial Neural Network outperformed the other methods in both training and prediction of temperature.

2.3 Comparison between existing works

From Literature review, many variations are found in machine learning methods and algorithms among the reviewed research papers. The various research papers have found different results with the same machine learning approaches. A machine learning model shows high accuracy with a research paper while same machine learning model may show low accuracy in another research. A comparative analysis with previous work is given in Table 2.1.

Table 2.1: Comparative Analysis Table

SL No	Author Name	Used Algorithm	Best Accuracy with Algorithm	Limitations
1.	Sashank, C. et al.	Linear Regression, Lasso Regression, Ridge Model, KNN Model, Random Forest Model	Random Forest Model = 42.2%	Data Dependency and Quality Issues, Geographical and Regional Variability, Complexity and Interpretability of the Model, Low Model Accuracy
2.	Gujnatti Rudrappa et al	Logistic Regression, Decision Tree, Neural Network, Random Forest, Light GBM, Catboost, XGBoost	XGBoost = 95%	Dependency on Data Quality and Pre-Processing, Variability in Model Performance Across Different Algorithms, Challenges in Generalizing Predictions to Different Locations
3.	Kumar, V. et al.	Random Forest, KNN (n=27), Decision Tree, Logistic Regression	Random Forest = 88.21%	Dependency on Data Quality and Completeness, Generalization Across Different Geographical Locations, Comparative Performance of Algorithms
4.	Shri Sandeep et al.	Random Forest, Artificial Neural Networks, Logistic	Random Forest = 87.76%	Complexity and Accuracy of Modular Models, Dependence

SL No	Author Name	Used Algorithm	Best Accuracy with Algorithm	Limitations
		Regression, Naïve Bayes model		on Historical Data Quality and Relevance, Challenges in Predicting Long-Term Rainfall
5.	Singla, P. et al.	Logistic Regression, Decision Tree, Random Forest, XGBoost	XGBoost = 85.73%	Data Quality and Relevance, Model Performance Variability, Generalization Across Regions
6.	Poornima, S. et al.	Intensified LSTM, LSTM, RNN with silu, RNN with relu, ELM, ARIMA, Holt-WInters	Intensified LSTM = 88%	Complexity and Computational Cost, Dependence on Historical Data Quality, Handling Extreme Events and Nonlinearity
7.	Rahman, A. et al.	Support Vector Machine (SVM), K-nearest Neighbors (KNN), Decision Tree, Naïve Bayes	Support Vector Machine (SVM) = 93%	Dependence on Historical Data Quality and Relevance, Complexity of Integration and Computational Overhead, Real-time Data Processing and Adaptability
8.	Bindu, B. H.	Naïve Bayes, Logistic Regression, ANN, Random Forest	Random Forest = 87.76%	Data Quality and Quantity, Model Complexity and Selection, Real-Time Data Processing and Adaptability
9.	Vishwakarma, H.	Linear Regression (LR), Support Vector Regression (SVR), lasso, ElasticNet	CO_2 Emission: Linear Regression = 98.52%, CH_4 Emission: Linear Regression = 93.66%, N_2O Emission: Linear Regression = 99.43%, SF_6	Model Limitations and Assumptions, Data Quality and Feature Selection, Generalization and Predictive Accuracy

SL No	Author Name	Used Algorithm	Best Accuracy with Algorithm	Limitations
			Emission: Linear Regression = 99.12%	
10.	Azari, B.	Linear Regression (LR), k-Nearest Neighbor (kNN), Support Vector Machine (SVM), Artificial Neural Network (ANN), Random Forest (RF), Adaptive Boosting (AdB)	Artificial Neural Network = 89.56%	Model Generalization and Complexity, Data Dependency and Feature Selection, Performance Evaluation and Benchmarking

2.4 Open Issues

The literature review reveals several open issues and gaps in the current research on predicting future changes in seasonal rainfall and temperature using machine learning models. These open issues include:

Scope Limitation:

- Many studies focus solely on rainfall prediction without considering temperature, greenhouse gases, or other climatic factors (e.g., Mekanik et al. [3], Alizadeh et al. [11]).
- Some studies are region-specific and do not address the generalizability of their findings to other regions or climates (e.g., Oswal et al. [4], Byun et al. [12]).

Model Limitations and Comparisons:

- There is a lack of detailed discussion on the limitations and challenges of specific machine learning models used for climate predictions (e.g., Sashank et al. [1], Rudrappa et al. [2]).
- Studies often present varying results with the same machine learning models, indicating inconsistencies in model performance across different datasets and regions (e.g., Kumar et al. [6], Shri Sandeep et al. [9]).

Data Quality and Pre-processing:

- Issues related to the quality, granularity, and completeness of historical climate data are not thoroughly addressed, impacting the accuracy and reliability of predictions (e.g., Liew et al. [8], Barrera-Animas et al. [15]).
- The impact of data pre-processing techniques on model performance is often underexplored (e.g., Oswal et al. [4]).

Long-term vs. Short-term Predictions:

- Many studies focus on short-term predictions, which may not be directly applicable to long-term climate forecasting (e.g., Sashank et al. [1], Rudrappa et al. [2]).
- Long-term predictions remain challenging due to the complex and dynamic nature of climatic variables (e.g., Kumar et al. [6]).

Integration of Machine Learning Techniques:

- Few studies explore the integration of different machine learning techniques, such as combining neural networks with optimization algorithms, to improve prediction accuracy (e.g., Azad et al. [5], Poornima et al. [13]).
- The potential of transfer learning and ensemble methods to enhance model performance is not fully exploited (e.g., Rahman et al. [14], Barrera-Animas et al. [15]).

Ethical Considerations and Transparency:

- Ethical considerations related to responsible data handling and transparent communication of findings are not sufficiently discussed (general issue across studies).
- The potential biases in data and model outputs and their implications for decision-making in climate resilience are not adequately addressed.

Sustainability and Practical Applications:

- The practical applications of predictions for adaptive agricultural practices, water resource management, and public health initiatives are often mentioned but not deeply explored (e.g., Vishwakarma et al. [23], Oyana et al. [7]).

- Studies lack a comprehensive approach to linking predictions with actionable strategies for enhancing climate resilience and sustainability (e.g., Bindhu et al. [22], Joseph et al. [19]).

Generalizability and Overfitting:

- The generalizability of machine learning models to diverse geographic areas and different time frames is not adequately validated (e.g., Wang et al. [21], Azari et al. [24]).
- The risk of overfitting, particularly with complex models like neural networks, is not sufficiently addressed, which can limit the reliability of predictions (e.g., Dutta et al. [16], Salcedo-Sanz et al. [17]).

Addressing these open issues requires a comprehensive and systematic approach that includes robust data collection and preprocessing, thorough validation of models across diverse regions and time frames, and integration of advanced machine learning techniques. Additionally, incorporating ethical considerations, transparency, and practical applications into the research will enhance the overall impact and relevance of climate predictions.

2.5 Summary

The literature review offers a thorough exploration into the forecasting of future changes in seasonal rainfall and temperature in Bangladesh, highlighting the application of different machine learning models. It introduces with a basic insight of climate science and investigates diverse studies utilizing machine learning in weather modeling. Remarkable works comprise a rainfall forecasting tool using neural networks, a relationship of machine learning algorithms for rainfall forecasting, and inventive methods such as using cloud image data for rainfall prediction. However, the review recognizes significant gaps, including a dearth of methodical comparative analyses, standardization of methodologies, and definite arguments on model limits and generalizability. The different findings from different studies using the same machine learning models emphasize the need for a more standardized and comprehensive method. The review advises that forthcoming research should center on adopting these gaps to boost up the strength and reliability of machine learning models in forecasting seasonal climate parameters for Bangladesh.

CHAPTER 3
Methodology

3.1 Overview

The methodology section of this thesis forms a comprehensive method to explore and predict future changes in seasonal rainfall and temperature in Bangladesh. The research design includes an organized collection of historical rainfall and temperature data of Bangladesh. A key focus is on machine learning models, including Linear Regression Model, Polynomial Regression Model, Decision Trees, Random Forest, Adaboost, Gradient Boosting Machines, employed to the dataset for rainfall prediction. The application of these models is expected to encapsulate intricate relationships and patterns essential in weather systems. The section underlines the importance of data preprocessing, model training, and assessment using suitable system of measurement like precision, mean square error and coverage. Moreover, the methodology investigates possible difficulties, limits and ambiguities conjoined with the chosen models, paying to a clear understanding of the research approach. In general, the methodology is organized to arrange a vigorous outline for the systematic investigation and forecasting of seasonal weather variables such as rainfall and temperature in Bangladesh.

3.2 Proposed Methodology

The methodology section of a research study serves as a blueprint that outlines how the research objectives will be achieved and how the data will be analyzed to answer the research questions. It provides a clear and systematic approach to conducting the study, ensuring transparency and reproducibility of the research process. In the context of predicting future changes in seasonal rainfall and temperature in Bangladesh using machine learning models, the methodology introduces the framework within which the research is conducted, encompassing data collection, processing, analysis techniques, and model development.

Proposed System Design:

The proposed system design in the methodology part revolves around integrating advanced machine learning models with historical climate data to predict future changes in seasonal rainfall and temperature patterns in Bangladesh. This design includes data preprocessing steps such as data extraction, cleaning, and feature engineering using tools like pandas and numpy. Machine learning algorithms such as linear regression, polynomial regression, decision tree, random forest, adaboost and gradient boosting machines are implemented using libraries such as scikit-learn within a Python environment like Jupyter Notebook. The engineering design diagram of whole research system is given in Figure 3.1.

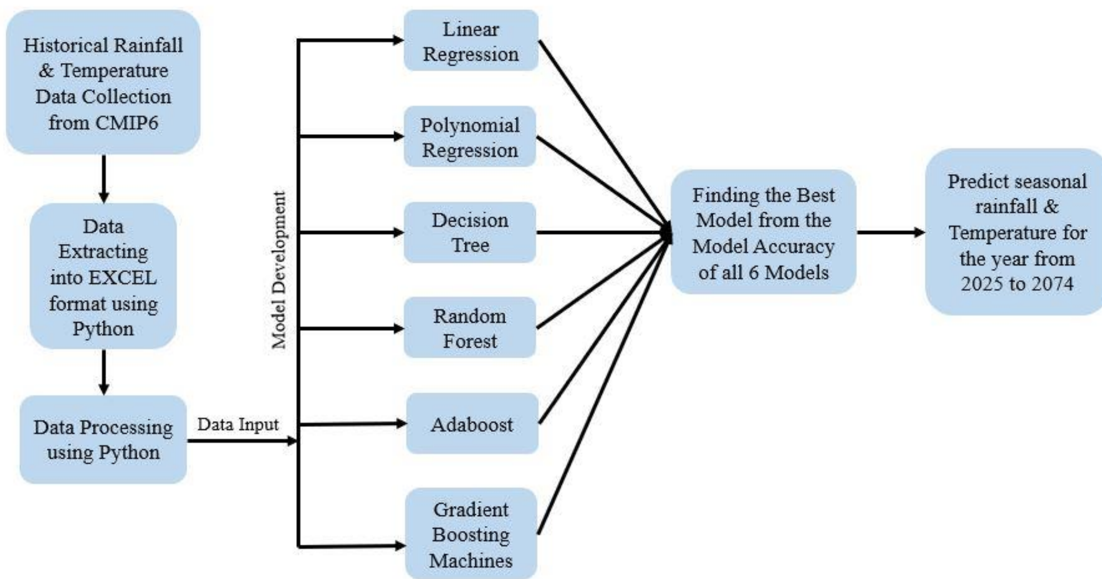

Figure 3.1: Engineering Design Diagram of Whole Research System

Data Collection and Preprocessing:

The methodology begins with the collection of historical climate data from 188 locations across Bangladesh. These datasets include daily records of rainfall, daily maximum temperature, and daily minimum temperature spanning the period from 1951 to 2014, sourced from the Coupled Model Intercomparison Project Phase 6 (CMIP6). The collected data is subjected to rigorous preprocessing steps to ensure quality and consistency. This involves data cleaning to handle missing values, outlier detection, and normalization to standardize the data for further analysis. Python libraries such as pandas and numpy are utilized for efficient data manipulation and preprocessing tasks.

Datasets:

Historical rainfall and temperature dataset (1951-2014) (CMIP6) for Bangladesh were collected from Intergovernmental Panel on Climate Change (IPCC) website. CMIP6 stands for Coupled Model Intercomparison Project Phase 6. The dataset was officially verified by professional research-based organization Institute of Water Modelling (IWM). The data was not in compatible format such as excel, csv, text files etc. So, python codes were used to extract the data into a compatible excel format. Python codes for extracting data are given below:

Location of Rainfall & Temperature Data:

The datasets used in this study are derived from historical climate data collected from 188 distinct locations across Bangladesh, spanning from 1951 to 2014. These datasets are sourced from the Coupled Model Intercomparison Project Phase 6 (CMIP6), which provides comprehensive daily records of meteorological parameters. The primary data points include daily rainfall, daily maximum temperature, and daily minimum temperature. The location of historical rainfall and temperature data are shown in Figure 3.2.

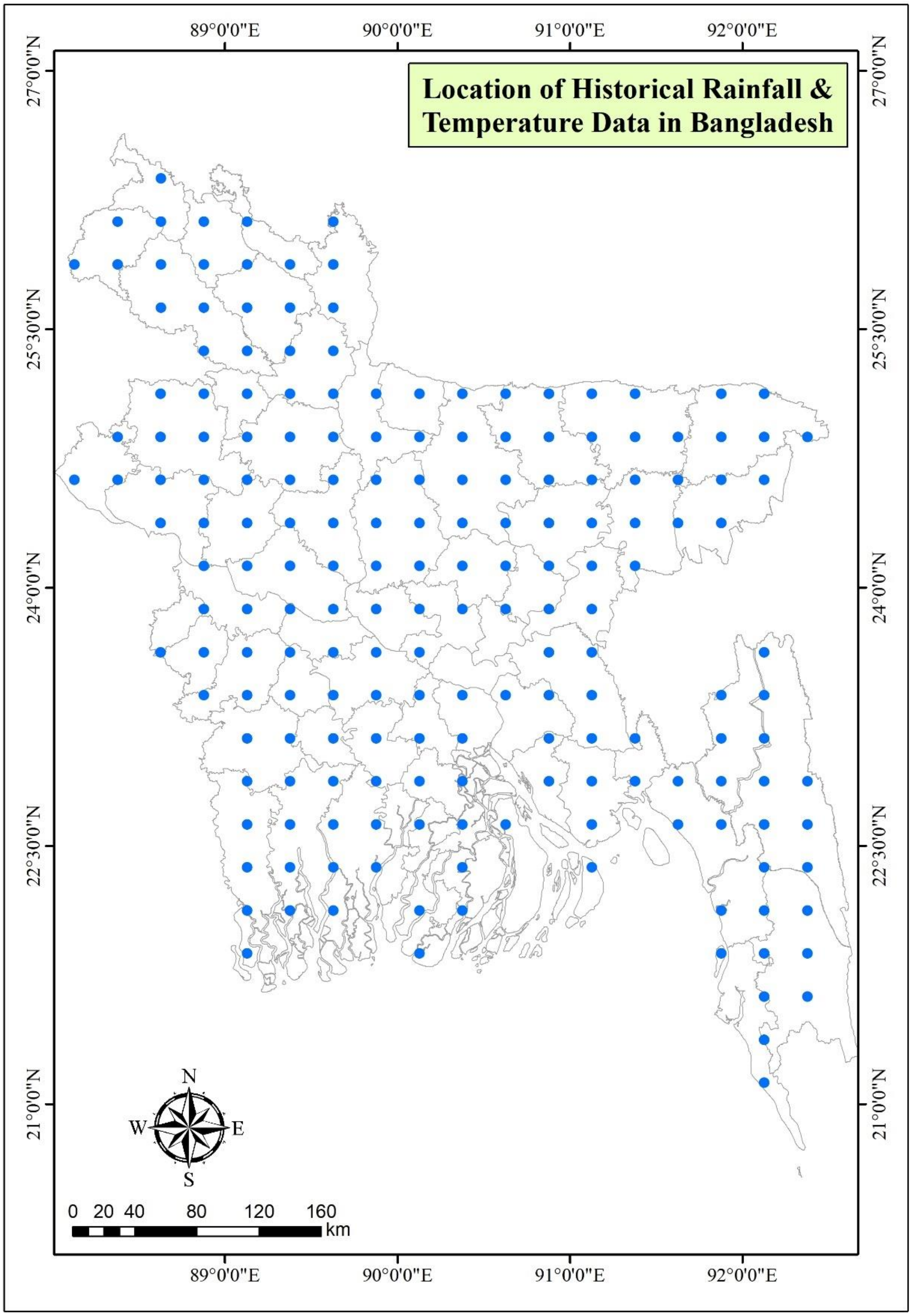

Figure 3.2: Location of historical rainfall and temperature data

Data Extraction & Preprocessing:

Rainfall Data Extraction:

```
import pandas as pd
import datetime as dt
import numpy as np

df=pd.read_table(r'F:\Thesis\data\Bangladesh\historical\PrecipData', sep=' ',header=None)
row1=list(df.iloc[0,3:])
row2=list(df.iloc[1,3:])
new_row=["Year","Month","Day"]
for r1,r2 in zip(row1,row2):
    values="({},{})".format(r1,r2)
    new_row.append(values)
df.columns=new_row
df.drop([0,1],axis=0,inplace=True)
df.reset_index(drop=True,inplace=True)

df.to_excel(r"F:\Thesis\data\Bangladesh\processed\historical_rainfall.xlsx")
```

Figure 3.3: Daily Rainfall data extraction using Python

Daily Maximum Temperature Data Extraction:

```
import pandas as pd
import datetime as dt
import numpy as np

df=pd.read_table(r'F:\Thesis\data\Bangladesh\historical\TMaxData', sep=' ',header=None)
row1=list(df.iloc[0,3:])
row2=list(df.iloc[1,3:])
new_row=["Year","Month","Day"]
for r1,r2 in zip(row1,row2):
    values="({},{})".format(r1,r2)
    new_row.append(values)
df.columns=new_row
df.drop([0,1],axis=0,inplace=True)
df.reset_index(drop=True,inplace=True)

df.to_excel(r"F:\Thesis\data\Bangladesh\processed\historical_TMaxData_data.xlsx")
```

Figure 3.4: Daily Maximum Temperature data extraction using Python

Daily Minimum Temperature Data Extraction:

```
import pandas as pd
import datetime as dt
import numpy as np

df=pd.read_table(r'F:\Thesis\data\Bangladesh\historical\TMinData', sep=' ',header=None)
row1=list(df.iloc[0,3:])
row2=list(df.iloc[1,3:])
new_row=["Year","Month","Day"]
for r1,r2 in zip(row1,row2):
    values="({},{})".format(r1,r2)
    new_row.append(values)
df.columns=new_row
df.drop([0,1],axis=0,inplace=True)
df.reset_index(drop=True,inplace=True)

df.to_excel(r"F:\Thesis\data\Bangladesh\processed\historical_TMinData_data.xlsx")
```

Figure 3.5: Daily Minimum Temperature data extraction using Python

Model Implementation:

Following data preprocessing, the thesis employs six machine learning models to forecast future climate variables:

- **Linear Regression:** This model establishes a linear relationship between the input features (such as historical climate data) and the target variable (future rainfall or temperature). It assumes a linear approach to predict outcomes based on weighted input features.
- **Polynomial Regression:** Extending linear regression, polynomial regression fits a polynomial equation to the data, allowing for more complex relationships between variables. This model captures non-linear patterns that may exist in the climate data.
- **Decision Tree:** Decision trees recursively partition the data into subsets based on the most significant features, creating a tree-like structure. Each internal node represents a "decision" based on a feature, leading to leaf nodes that provide the predicted output.
- **Random Forest:** A robust ensemble learning method, random forest constructs multiple decision trees during training and outputs the average prediction of individual trees. This ensemble approach reduces overfitting and enhances prediction accuracy.

- **Adaboost:** AdaBoost (Adaptive Boosting) is an ensemble learning method that combines multiple weak learners, typically decision trees, to create a strong predictive model by iteratively improving on the errors of previous models. It adjusts the weights of misclassified instances in each iteration to focus learning on harder cases, thereby enhancing overall accuracy and robustness.
- **Gradient Boosting Machines:** Gradient Boosting Machines (GBM) is an ensemble learning technique that builds a series of decision trees sequentially, where each tree corrects the errors of its predecessor by fitting to the residuals of the combined ensemble. This iterative process optimizes a loss function, leading to a robust model that improves predictive performance and handles complex patterns in data.

Model Training and Evaluation:

Each machine learning model is trained on a portion of the preprocessed dataset using appropriate training techniques. Cross-validation methods such as k-fold validation ensure that the models generalize well to unseen data and mitigate overfitting. The training process aims to optimize model parameters to improve accuracy in predicting future climate variables.

Performance Evaluation:

The performance of each model is evaluated using R-squared value (coefficient of determination). The R-squared value measured by the model assesses the model's ability to accurately predict seasonal rainfall and temperature changes compared to actual historical data. Comparative analysis across the six models identifies the most effective approach for forecasting climate variables in Bangladesh.

In summary, the proposed methodology integrates comprehensive data preprocessing, model implementation using six distinct machine learning techniques, rigorous model training and evaluation, and comparative performance analysis. This structured approach aims to deliver reliable predictions of future changes in seasonal climate patterns, supporting informed decision-making in sectors crucially impacted by climate variability in Bangladesh.

Prediction

Once the best-performing model is identified, it is used to predict future changes in seasonal rainfall and average temperature. The prediction phase involves inputting new, unseen data into the model to generate forecasts. These predictions provide valuable insights into potential future climatic conditions, which are crucial for planning and mitigation strategies in various sectors, such as agriculture, water resource management, and public health. The ability to accurately forecast climatic changes enables stakeholders to make informed decisions to enhance climate resilience.

Analysis and Interpretation

The final step in the experimental process is the analysis and interpretation of the predictions. This involves examining the predicted values to understand the potential trends and patterns in future seasonal rainfall and temperature. The study interprets these results in the context of Bangladesh's specific climatic and geographical conditions, providing actionable insights for policymakers and other stakeholders. The findings are then communicated transparently, highlighting the strengths and limitations of the models used, and suggesting areas for future research to further improve predictive accuracy and applicability.

By following this structured and rigorous experimental process, the study aims to contribute significantly to the understanding and forecasting of climatic changes in Bangladesh, ultimately aiding in the development of effective strategies to mitigate the impacts of climate change.

Results:

The model results contain the future prediction of seasonal rainfalls and average temperature. The predicted values for the year from 2025 to 2074 are shown in tabular format. Moreover, the percentage of change of seasonal rainfall and average temperature with respect to 2024 are also presented in the table. The comparisons of percentage of seasonal rainfall and average temperature in degree Celsius between 2024 and 2050 are shown in a graph.

Documentation and Reporting:

Documenting the entire implementation process, including data preprocessing steps, model configurations, parameter tuning results, and performance metrics. This facilitates reproducibility of results and provides a comprehensive basis for reporting findings in the thesis or research paper.

3.3 Hardware/Software Requirement

Implementing the proposed methodology for predicting future changes in seasonal rainfall and temperature in Bangladesh using machine learning models involves several key requirements of hardware and software:

- Hardware: A robust computer or server with sufficient processing power (CPU) and memory (RAM) to handle large datasets and complex computations. Depending on the size of the dataset and the complexity of the models, a high-performance workstation or cloud computing resources may be necessary.
- Software: Installation of appropriate software packages and libraries for data preprocessing, model implementation, and evaluation. This thesis typically involves Python as the primary programming language due to its extensive libraries such as pandas for data manipulation, numpy for numerical operations, and scikit-learn (sklearn) for machine learning algorithms. Moreover, the study involves ArcGIS software for showing the data location through map.

Programming Environment:

Setting up a conducive programming environment is crucial for developing and testing machine learning models:

- Integrated Development Environment (IDE): Using IDEs like Jupyter Notebook or Spyder for interactive development and experimentation with code.
- Version Control: Employing version control systems (e.g., Git) to manage code changes, collaborate with team members (if applicable), and track experimental setups.

3.4 Project Management & Financial Analysis

Managing the execution of this thesis involves a systematic approach to ensure the successful completion of research objectives within defined timelines and resource constraints. The project management framework encompasses several key components:

Project Initiation:

Objective: Define the scope, objectives, and stakeholders of the thesis.

Components:

- Develop a project charter outlining the purpose, goals, and stakeholders
- Clearly define the research questions and expected outcomes

Planning and Scheduling:

Objective: Establish a detailed plan for executing the research

Components:

- Create a comprehensive project plan with timelines for each phase
- Identify and allocate resources, including data sources and computing resources
- Define milestones for key deliverables such as literature review, methodology, and results

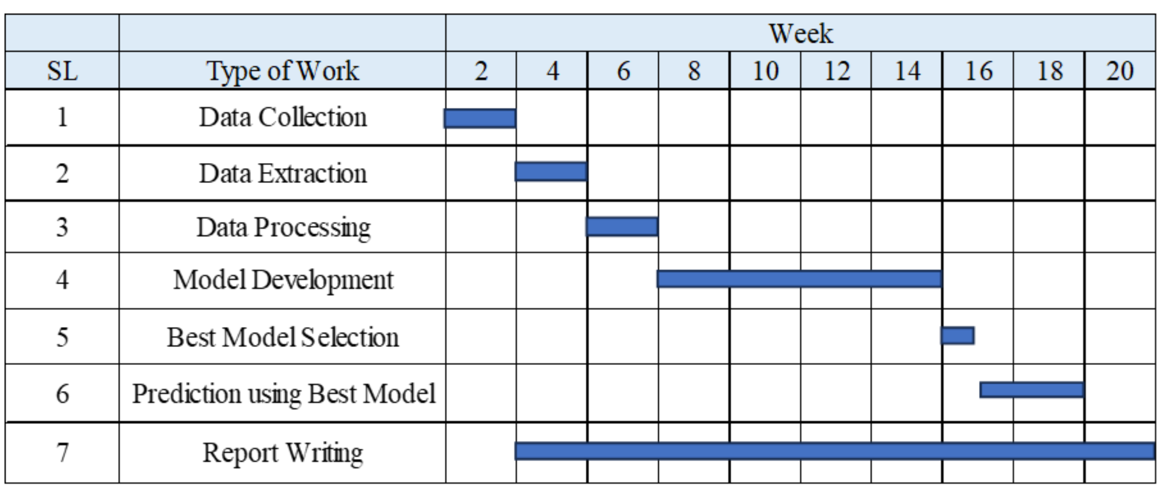

Figure 3.6: Activity Schedule of the Research Study

Risk Management:

Objective: Identify and mitigate potential risks to the project

Components:

- Conduct a risk assessment to identify potential challenges in data collection, model implementation, or analysis.
- Develop mitigation strategies for identified risks.

Stakeholder Communication:

Objective: Ensure transparent and effective communication with stakeholders

Components:

- Establish regular communication channels with advisors, peers, and other relevant stakeholders
- Provide updates on project progress, challenges, and successes

Resource Management:

Objective: Efficiently allocate and utilize available resources

Components:

- Ensure access to necessary data sources and computational resources.
- Monitor resource usage to prevent bottlenecks or delays

Iterative Development:

Objective: Embrace an iterative approach to project development

Components:

- Allow for flexibility in the research plan to accommodate insights gained during the process.
- Incorporate feedback from advisors and peers for continuous improvement.

Documentation and Reporting:

Objective: Maintain comprehensive documentation and reporting practices

Components:

- Document all phases of the research, including data collection, preprocessing steps, model implementations, and results.
- Prepare regular progress reports and share findings with stakeholders.

Evaluation and Reflection:

Objective: Assess project progress and reflect on outcomes

Components:

- Regularly evaluate the effectiveness of the chosen methodologies
- Reflect on challenges faced and lessons learned for continuous improvement

This project management approach aims to provide a structured and adaptive framework, allowing for efficient collaboration, risk mitigation, and continuous improvement throughout the lifecycle of the thesis. Regular monitoring and adaptation of the plan based on ongoing assessments contribute to the overall success of the research endeavor.

Financial Analysis

The CMIP6 dataset is an open-source data and fully free. So, there was no need to buy the dataset. But the dataset has been verified by a professional organization Institute of Water Modelling (IWM).

Moreover, there is no cost to conduct the whole thesis because the thesis is based on machine learning models which can be done completely in python programming and python programming

g interface is completely free and open source.

The estimated cost for the study work is shown below in the Table 3.1.

Table 3.1: Estimated Cost for Machine Learning based project

SN	Components	Estimated Costs (BDT)
1	Visiting Stakeholders	1500-2000
2	Data Verification	500-1000
3	Miscellaneous	200-300
	Total Estimated Cost	2200-3300

3.5 Summary

The methodology for predicting future changes in seasonal rainfall and temperature in Bangladesh involves a structured approach starting with the collection and preprocessing of historical climate data from 188 locations. This data, spanning from 1951 to 2014, is sourced from the CMIP6 and verified by the Institute of Water Modelling (IWM). Six machine learning models - Linear Regression, Polynomial Regression, Decision Trees, Random Forest, Adaboost and Gradient Boosting Machines - are employed to predict climate variables. These models undergo rigorous training and evaluation, with performance assessed using metrics like R-squared value, Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). The methodology includes prediction phases for generating future climate forecasts, which are analyzed and interpreted to provide actionable insights for stakeholders. Documentation and reporting are emphasized to ensure transparency and reproducibility, with a detailed project management plan in place to guide the research. Finally, the financial analysis highlights minimal costs due to the open-source nature of the dataset and Python programming.

CHAPTER 4

Implementation

4.1 Overview

The implementation of this thesis focuses on utilizing advanced machine learning models to predict future changes in seasonal rainfall and temperature across Bangladesh. Beginning with the collection and rigorous preprocessing of historical climate data from 1951 to 2014, sourced from 188 locations and verified by the CMIP6 dataset and Institute of Water Modelling (IWM), the study ensures data quality and consistency. Machine learning models including Linear Regression, Polynomial Regression, Decision Trees, Random Forest, Adaboost and Gradient Boosting Machines are then employed to analyze the preprocessed data and forecast climate variables. The models undergo thorough training, parameter optimization, and evaluation using metrics like R-squared value, RMSE and MAE to ensure accuracy in predicting climatic patterns. Results are interpreted to provide actionable insights for stakeholders in sectors such as agriculture and water management, supported by comprehensive documentation and adherence to project management practices for transparency and reproducibility.

4.2 Train Model

Six machine learning models are developed in this study. They are linear regression, polynomial regression, decision tree, random forest, adaboost and gradient boosting machines. The raw data is processed for preparing it to incorporate into models. Then six machine learning models are setup using python programming. Then model performances of the models are measured and compared using model accuracy. Based on the model accuracy a machine learning model is selected for predicting future seasonal rainfall and temperature for next 50 years (2025-2074). Moreover, percentage of change of seasonal rainfall and change of seasonal average temperature for future year comparing with 2024 are determined.

There are 188 locations of Bangladesh having historical time series data of CMIP6 such as daily rainfall, daily maximum temperature and daily minimum temperature data for 1951 to 2014. Then all daily rainfall data are converted into a single time series data for whole Bangladesh averaging the data of 188 locations. Similar conversion is done for daily maximum temperature data and daily minimum data and then two time series data of whole Bangladesh for daily maximum and daily minimum data are converting into single daily average temperature time series for whole Bangladesh averaging daily maximum and daily minimum temperature data. Now two time series files are created: one is daily rainfall time series and the other is daily average temperature time series representing whole Bangladesh.

The two time series data such as rainfall and average temperature time series for 1951 to 2014) are categorized into four seasons: Premonsoon (March, April, May), Monsoon (June, July, August, September), Postmonsoon (October, November) and Winter (December, January, February). Then finally eight types of data are prepared for model input:

- Yearly sum of Premonsoon rainfall (1951-2014)
- Yearly sum of Monsoon rainfall (1951-2014)
- Yearly sum of Postmonsoon rainfall (1951-2014)
- Yearly sum of Winter rainfall (1951-2014)
- Yearly average of Premonsoon temperature (1951-2014)
- Yearly average of Monsoon temperature (1951-2014)
- Yearly average of Postmonsoon temperature (1951-2014)
- Yearly average of Winter temperature (1951-2014)

For every type of data there are six machine learning model setups are prepared which results in 48 distinct machine learning models. Model setups included training and testing the data and making the models suitable for necessary prediction. Then only one machine learning model setup is selected for every distinct type of yearly series of data based on model accuracy.

Python codes of training and testing of data as well as predicting the future seasonal rainfall and temperature data for selected models of eight types of data are given below:

Yearly sum of Premonsoon rainfall

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score

df = pd.read_excel(r'F:\Thesis\predefense\model\rainfall\input\premonsoon.xlsx')
X = df['Year']
X = np.array(X)
y = df['Rainfall (mm)']
y = np.array(y)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

X_train = X_train.reshape(-1, 1)
X_test = X_test.reshape(-1, 1)
y_train = y_train.reshape(-1, 1)
y_test = y_test.reshape(-1, 1)

poly_features = PolynomialFeatures(degree=3)
X_train_poly = poly_features.fit_transform(X_train)
X_test_poly = poly_features.transform(X_test)

model = LinearRegression()
model.fit(X_train_poly, y_train)
y_train_pred = model.predict(X_train_poly)
y_test_pred = model.predict(X_test_poly)
r2 = r2_score(y_test, y_test_pred)
efficiency_percentage = r2 * 100
print("Model Efficiency:", efficiency_percentage, "%")
DF = pd.DataFrame()
years = np.arange(2024, 2074)
```

```
predicted = [ ]
for year in years:
    X_train_poly = poly_features.fit_transform([[year]])
    y_pred = model.predict(X_train_poly)
    y_pred = y_pred.item()
    y_pred = round(y_pred)
    predicted.append(y_pred)
DF["Year"] = years
DF["Predcited Premonsoon Rainfall (mm)"] = predicted
DF.to_excel(r'F:\Thesis\predefense\model\rainfall\output\premonsoon.xlsx', index =
False)
```

<u>Yearly sum of Monsoon rainfall</u>

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score

df = pd.read_excel(r'F:\Thesis\predefense\model\rainfall\input\monsoon.xlsx')
X = df['Year']
X = np.array(X)
y = df['Rainfall (mm)']
y = np.array(y)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

X_train = X_train.reshape(-1, 1)
X_test = X_test.reshape(-1, 1)
y_train = y_train.reshape(-1, 1)
y_test = y_test.reshape(-1, 1)
```

```
poly_features = PolynomialFeatures(degree=2)
X_train_poly = poly_features.fit_transform(X_train)
X_test_poly = poly_features.transform(X_test)

model = LinearRegression()
model.fit(X_train_poly, y_train)

y_train_pred = model.predict(X_train_poly)
y_test_pred = model.predict(X_test_poly)

r2 = r2_score(y_test, y_test_pred)
efficiency_percentage = r2 * 100
print("Model Efficiency:", efficiency_percentage, "%")

DF = pd.DataFrame()
years = np.arange(2024, 2074)
predicted = [ ]
for year in years:
    X_train_poly = poly_features.fit_transform([[year]])
    y_pred = model.predict(X_train_poly)
    y_pred = y_pred.item()
    y_pred = round(y_pred)
    predicted.append(y_pred)
DF["Year"] = years
DF["Predicted Monsoon Rainfall (mm)"] = predicted
DF.to_excel(r'F:\Thesis\predefense\model\rainfall\output\monsoon.xlsx', index = False)
```

Yearly sum of Postmonsoon Rainfall

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import PolynomialFeatures
```

```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score

df = pd.read_excel(r'F:\Thesis\predefense\model\rainfall\input\postmonsoon.xlsx')
X = df['Year']
X = np.array(X)
y = df['Rainfall (mm)']
y = np.array(y)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

X_train = X_train.reshape(-1, 1)
X_test = X_test.reshape(-1, 1)
y_train = y_train.reshape(-1, 1)
y_test = y_test.reshape(-1, 1)

poly_features = PolynomialFeatures(degree=1)
X_train_poly = poly_features.fit_transform(X_train)
X_test_poly = poly_features.transform(X_test)

model = LinearRegression()
model.fit(X_train_poly, y_train)

y_train_pred = model.predict(X_train_poly)
y_test_pred = model.predict(X_test_poly)

r2 = r2_score(y_test, y_test_pred)
efficiency_percentage = r2 * 100
print("Model Efficiency:", efficiency_percentage, "%")
DF = pd.DataFrame()
years = np.arange(2024, 2074)
predicted = [ ]
for year in years:
    X_train_poly = poly_features.fit_transform([[year]])
```

```
    y_pred = model.predict(X_train_poly)
    y_pred = y_pred.item()
    y_pred = round(y_pred)
    predicted.append(y_pred)
DF["Year"] = years
DF["Predcited Postmonsoon Rainfall (mm)"] = predicted
DF.to_excel(r'F:\Thesis\predefense\model\rainfall\output\postmonsoon.xlsx', index =
False)
```

Yearly Sum of Winter Rainfall

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score

df = pd.read_excel(r'F:\Thesis\predefense\model\rainfall\input\winter.xlsx')
X = df['Year']
X = np.array(X)
y = df['Rainfall (mm)']
y = np.array(y)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

X_train = X_train.reshape(-1, 1)
X_test = X_test.reshape(-1, 1)
y_train = y_train.reshape(-1, 1)
y_test = y_test.reshape(-1, 1)
poly_features = PolynomialFeatures(degree=1)
X_train_poly = poly_features.fit_transform(X_train)
X_test_poly = poly_features.transform(X_test)
```

```
model = LinearRegression()
model.fit(X_train_poly, y_train)

y_train_pred = model.predict(X_train_poly)
y_test_pred = model.predict(X_test_poly)

r2 = r2_score(y_test, y_test_pred)
efficiency_percentage = r2 * 100
print("Model Efficiency:", efficiency_percentage, "%")

DF = pd.DataFrame()
years = np.arange(2024, 2074)
predicted = [ ]
for year in years:
    X_train_poly = poly_features.fit_transform([[year]])
    y_pred = model.predict(X_train_poly)
    y_pred = y_pred.item()
    y_pred = round(y_pred)
    predicted.append(y_pred)
DF["Year"] = years
DF["Predcited Winter Rainfall (mm)"] = predicted
DF.to_excel(r'F:\Thesis\predefense\model\rainfall\output\winter.xlsx', index = False)
```

Yearly average of Premonsoon temperature

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score
```

```
df = pd.read_excel(r'F:\Thesis\predefense\model\temperature\input\premonsoon.xlsx')
X = df['Year']
X = np.array(X)
y = df['Temp (Degree C)']
y = np.array(y)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

X_train = X_train.reshape(-1, 1)
X_test = X_test.reshape(-1, 1)
y_train = y_train.reshape(-1, 1)
y_test = y_test.reshape(-1, 1)

poly_features = PolynomialFeatures(degree=1)
X_train_poly = poly_features.fit_transform(X_train)
X_test_poly = poly_features.transform(X_test)

model = LinearRegression()
model.fit(X_train_poly, y_train)

y_train_pred = model.predict(X_train_poly)
y_test_pred = model.predict(X_test_poly)

r2 = r2_score(y_test, y_test_pred)
efficiency_percentage = r2 * 100
print("Model Efficiency:", efficiency_percentage, "%")

DF = pd.DataFrame()
years = np.arange(2024, 2074)
predicted = [ ]
for year in years:
    X_train_poly = poly_features.fit_transform([[year]])
    y_pred = model.predict(X_train_poly)
```

```
    y_pred = y_pred.item()
    predicted.append(y_pred)
DF["Year"] = years
DF["Predcited Premonsoon Temperature (Degree C)"] = predicted
DF.to_excel(r'F:\Thesis\predefense\model\temperature\output\premonsoon.xlsx',  index  =
False)
```

Yearly average of Monsoon temperature

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score

df = pd.read_excel(r'F:\Thesis\predefense\model\temperature\input\premonsoon.xlsx')
X = df['Year']
X = np.array(X)
y = df['Temp (Degree C)']
y = np.array(y)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

X_train = X_train.reshape(-1, 1)
X_test = X_test.reshape(-1, 1)
y_train = y_train.reshape(-1, 1)
y_test = y_test.reshape(-1, 1)

poly_features = PolynomialFeatures(degree=1)
X_train_poly = poly_features.fit_transform(X_train)
X_test_poly = poly_features.transform(X_test)

model = LinearRegression()
model.fit(X_train_poly, y_train)
```

```
y_train_pred = model.predict(X_train_poly)
y_test_pred = model.predict(X_test_poly)

r2 = r2_score(y_test, y_test_pred)
efficiency_percentage = r2 * 100
print("Model Efficiency:", efficiency_percentage, "%")

DF = pd.DataFrame()
years = np.arange(2024, 2074)
predicted = [ ]
for year in years:
    X_train_poly = poly_features.fit_transform([[year]])
    y_pred = model.predict(X_train_poly)
    y_pred = y_pred.item()
    predicted.append(y_pred)
DF["Year"] = years
DF["Predcited Premonsoon Temperature (Degree C)"] = predicted
DF.to_excel(r'F:\Thesis\predefense\model\temperature\output\premonsoon.xlsx', index = False)
```

Yearly average of Postmonsoon temperature

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score

df = pd.read_excel(r'F:\Thesis\predefense\model\temperature\input\premonsoon.xlsx')
X = df['Year']
X = np.array(X)
y = df['Temp (Degree C)']
```

```
y = np.array(y)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

X_train = X_train.reshape(-1, 1)
X_test = X_test.reshape(-1, 1)
y_train = y_train.reshape(-1, 1)
y_test = y_test.reshape(-1, 1)

poly_features = PolynomialFeatures(degree=1)
X_train_poly = poly_features.fit_transform(X_train)
X_test_poly = poly_features.transform(X_test)

model = LinearRegression()
model.fit(X_train_poly, y_train)

y_train_pred = model.predict(X_train_poly)
y_test_pred = model.predict(X_test_poly)

r2 = r2_score(y_test, y_test_pred)
efficiency_percentage = r2 * 100
print("Model Efficiency:", efficiency_percentage, "%")

DF = pd.DataFrame()
years = np.arange(2024, 2074)
predicted = [ ]
for year in years:
    X_train_poly = poly_features.fit_transform([[year]])
    y_pred = model.predict(X_train_poly)
    y_pred = y_pred.item()
    predicted.append(y_pred)
DF["Year"] = years
DF["Predcited Premonsoon Temperature (Degree C)"] = predicted
DF.to_excel(r'F:\Thesis\predefense\model\temperature\output\premonsoon.xlsx',  index  = False)
```

Yearly average of Winter temperature

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score

df = pd.read_excel(r'F:\Thesis\predefense\model\temperature\input\winter.xlsx')
X = df['Year']
X = np.array(X)
y = df['Temp (Degree C)']
y = np.array(y)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
X_train = X_train.reshape(-1, 1)
X_test = X_test.reshape(-1, 1)
y_train = y_train.reshape(-1, 1)
y_test = y_test.reshape(-1, 1)

poly_features = PolynomialFeatures(degree=1)
X_train_poly = poly_features.fit_transform(X_train)
X_test_poly = poly_features.transform(X_test)


model = LinearRegression()
model.fit(X_train_poly, y_train)
y_train_pred = model.predict(X_train_poly)
y_test_pred = model.predict(X_test_poly)

r2 = r2_score(y_test, y_test_pred)
efficiency_percentage = r2 * 100
print("Model Efficiency:", efficiency_percentage, "%")
```

```
DF = pd.DataFrame()
years = np.arange(2024, 2074)
predicted = [ ]
for year in years:
    X_train_poly = poly_features.fit_transform([[year]])
    y_pred = model.predict(X_train_poly)
    y_pred = y_pred.item()
    predicted.append(y_pred)
DF["Year"] = years
DF["Predcited Winter Temperature (Degree C)"] = predicted
DF.to_excel(r'F:\Thesis\predefense\model\temperature\output\winter.xlsx', index = False)
```

4.3 Model Evaluation

In this extensive evaluation, six machine learning models - Linear Regression, Polynomial Regression, Decision Tree, Random Forest, AdaBoost, and Gradient Boosting Machines - are assessed for their accuracy in predicting future seasonal rainfall and temperature data across eight different datasets. Using Python programming, the performance of each model is meticulously analyzed by calculating their accuracy metrics, which allows for a comprehensive comparison. This evaluation involves fitting each model to the data and measuring their predictive performance through metrics such as R-squared, RMSE, and MAE. The models demonstrating the highest accuracy and lowest error rates are identified and selected for future predictions. This methodical approach ensures that only the most reliable models are used for forecasting seasonal rainfall and temperature. The selected models are then employed to provide future climate predictions, offering valuable insights for climate research. These predictions can aid in strategic agricultural planning and resource management. By leveraging the strengths of various machine learning algorithms, this evaluation aims to enhance the accuracy and reliability of climate forecasts. Consequently, it supports informed decision-making in sectors reliant on weather and climate data.

4.4 Summary

The implementation of this thesis involves a rigorous process of utilizing machine learning models to predict future changes in seasonal rainfall and temperature across Bangladesh. Beginning with the collection and preprocessing of historical climate data from 1951 to 2014, sourced from 188 locations and validated by the CMIP6 dataset and Institute of Water Modelling (IWM), the study ensures data quality and consistency. Four types of yearly data are derived from this preprocessing, covering rainfall and average temperature across four seasons. Subsequently, 32 distinct machine learning models are developed and evaluated for each type of data, including Linear Regression, Polynomial Regression, Decision Trees, Random Forest, Adaboost and Gradient Boosting Machines. Model performance metrics, such as accuracy, guide the selection of the best-performing model for predicting future climate variables from 2025 to 2074. This approach culminates in actionable insights for stakeholders in agriculture, water management, and other sectors, supported by comprehensive documentation and adherence to rigorous scientific methodologies, ensuring transparency and reproducibility of results.

The model results are shown in a website named as *Weather Prediction House*. The website asks the expected year and season and gives the seasonal rainfall, percentage change of rainfall with respect to 2024, average temperature and change of temperature with respect to 2024 for the corresponding year and season. The images of the website are given below:

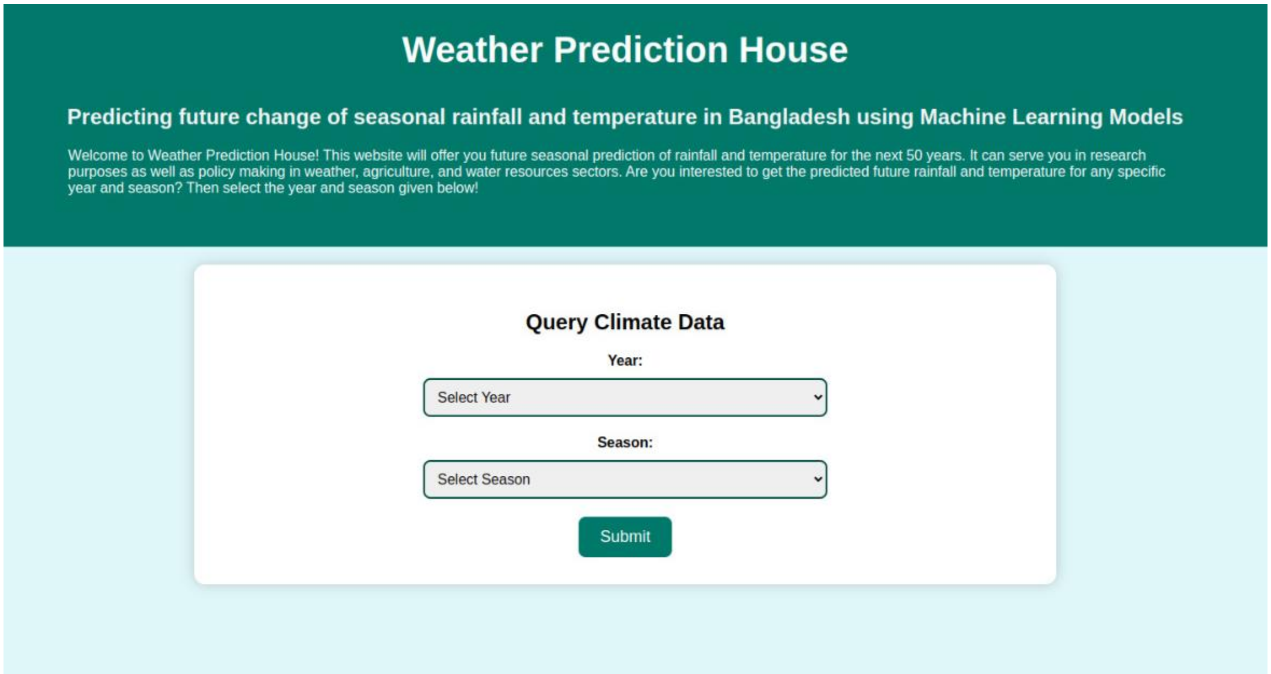

Figure 4.1: Creating Weather Prediction House Website

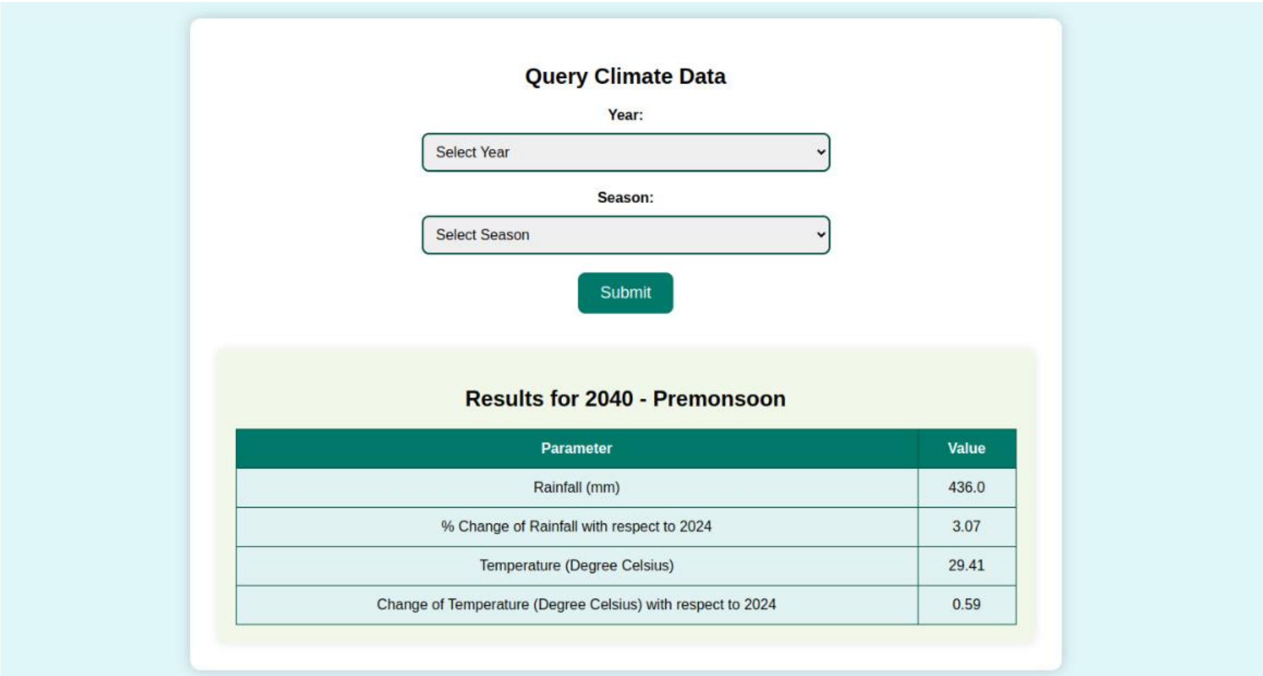

Figure 4.2: Model Results shown on Website for 50 years (2025-2074)

CHAPTER 5
Result & Analysis

5.1 Overview

The results of this thesis provide valuable insights into future climate trends in Bangladesh, based on predictions made using advanced machine learning models. Through rigorous evaluation and comparison of model performances, the study identifies the most accurate models for forecasting seasonal rainfall and temperature changes from 2025 to 2074. Results include predicted values and percentage changes compared to a baseline year, such as 2024, across different seasons. Analysis of these predictions highlights significant trends and potential implications for sectors reliant on climate-sensitive decision-making, such as agriculture and water resource management. The findings are presented with clarity, supported by statistical measures like R-squared values, Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) to enhance credibility and inform policy and adaptation strategies. Future research opportunities are also identified, aiming to improve model precision and expand the scope of climate forecasting capabilities in the region.

5.2 Simulation Result

Best Machine learning models are selected through comparing the model accuracy of six machine learning models for eight types of data. Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) are also determined from the models. The summary of the selection of the machine learning models for every yearly series of data are given in the Table 5.1. RMSE and MAE of the machine learning models are shown in Table 5.2 and Table 5.3 respectively. From Table 5.1, it is shown that for yearly sum of premonsoon rainfall and yearly sum of monsoon rainfall, polynomial regression model governs in respect of model accuracy and linear regression model governs for the rest of the types of times series data. Those models are also the best in respect of Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE).

Table 5.1: Summary of the selection of best machine learning model based on model accuracy

SL	Type of Data	Model Accuracy						Selected Model
		Linear Regression (%)	Polynomial Regression (%)	Decision Tree (%)	Random Forest (%)	Adaboost (%)	Gradient Boosting Machines (%)	
1	Yearly sum of Premonsoon rainfall	85.84	86.99	55.51	69.10	48.95	62.56	Polynomial Regression
2	Yearly sum of Monsoon rainfall	86.00	87.87	78.83	78.84	82.78	78.99	Polynomial Regression
3	Yearly sum of Postmonsoon rainfall	88.82	86.89	74.50	80.81	75.04	77.05	Linear Regression
4	Yearly sum of Winter rainfall	85.55	80.37	69.82	78.79	68.28	71.91	Linear Regression
5	Yearly average of Premonsoon temperature	88.63	87.84	82.53	83.55	79.55	82.37	Linear Regression
6	Yearly average of Monsoon temperature	90.44	88.97	49.44	75.29	60.54	67.62	Linear Regression
7	Yearly average of Postmonsoon temperature	88.80	86.98	81.62	86.88	81.44	82.39	Linear Regression
8	Yearly average of Winter temperature	87.47	87.17	69.07	86.00	69.76	79.81	Linear Regression

Table 5.2: Root Mean Squared Error (RMSE) of the models

SL	Type of Data	RMSE					
		Linear Regression	Polynomial Regression	Decision Tree	Random Forest	Adaboost	Gradient Boosting Machines
1	Yearly sum of Premonsoon rainfall	2.47	2.36	4.37	3.64	4.68	4.01
2	Yearly sum of Monsoon rainfall	89.84	83.65	110.48	110.47	99.66	110.06
3	Yearly sum of Postmonsoon rainfall	2.15	2.33	3.25	2.82	3.21	3.08
4	Yearly sum of Winter rainfall	1.07	1.25	1.55	1.3	1.58	1.49
5	Yearly average of Premonsoon temperature	0.30	0.31	0.38	0.37	0.41	0.38
6	Yearly average of Monsoon temperature	0.27	0.29	0.63	0.44	0.56	0.51
7	Yearly average of Postmonsoon temperature	0.30	0.32	0.38	0.32	0.38	0.37
8	Yearly average of Winter temperature	0.63	0.64	0.99	0.67	0.98	0.80

Table 5.3: Mean Absolute Error (MAE) of the models

SL	Type of Data	MAE					
		Linear Regression	Polynomial Regression	Decision Tree	Random Forest	Adaboost	Gradient Boosting Machines
1	Yearly sum of Premonsoon rainfall	1.72	1.71	3.34	2.78	3.91	2.97
2	Yearly sum of Monsoon rainfall	75.55	68.68	89.67	88.59	80.33	89.19
3	Yearly sum of Postmonsoon rainfall	1.50	1.70	2.67	2.30	2.62	2.54
4	Yearly sum of Winter rainfall	0.69	0.82	1.21	1.05	1.23	1.18
5	Yearly average of Premonsoon temperature	0.26	0.27	0.33	0.31	0.34	0.33
6	Yearly average of Monsoon temperature	0.19	0.20	0.54	0.35	0.46	0.41
7	Yearly average of Postmonsoon temperature	0.22	0.23	0.27	0.25	0.27	0.29
8	Yearly average of Winter temperature	0.42	0.42	0.91	0.54	0.89	0.73

The predicted values of future seasonal rainfall and average temperature for next 50 years (2025-2074) are extracted from the selected machine learning models. Then the percentage of rainfall change with respect to 2024 and the change of temperature are determined. The summary of the result of the prediction from machine learning models is given in Table 5.4.

Table 5.4: Predicted values of future seasonal rainfall and average temperature for next 50 years (2025-2074)

Year	**Season**	**Rainfall (mm)**	**% Change of Rainfall with respect to 2024**	**Temperature (Degree Celsius)**	**Change of Temperature (Degree Celsius) with respect to 2024**
2025	Premonsoon	423	0.00	28.86	0.04
2026	Premonsoon	424	0.24	28.89	0.07
2027	Premonsoon	425	0.47	28.93	0.11
2028	Premonsoon	426	0.71	28.97	0.15
2029	Premonsoon	426	0.71	29.00	0.18
2030	Premonsoon	427	0.95	29.04	0.22
2031	Premonsoon	428	1.18	29.08	0.26
2032	Premonsoon	429	1.42	29.11	0.29
2033	Premonsoon	429	1.42	29.15	0.33
2034	Premonsoon	430	1.65	29.19	0.37
2035	Premonsoon	431	1.89	29.23	0.41
2036	Premonsoon	432	2.13	29.26	0.44
2037	Premonsoon	433	2.36	29.30	0.48
2038	Premonsoon	434	2.60	29.34	0.52
2039	Premonsoon	435	2.84	29.37	0.55
2040	Premonsoon	436	3.07	29.41	0.59
2041	Premonsoon	437	3.31	29.45	0.63
2042	Premonsoon	438	3.55	29.48	0.66
2043	Premonsoon	439	3.78	29.52	0.70
2044	Premonsoon	440	4.02	29.56	0.74
2045	Premonsoon	441	4.26	29.59	0.77
2046	Premonsoon	442	4.49	29.63	0.81
2047	Premonsoon	443	4.73	29.67	0.85
2048	Premonsoon	444	4.96	29.71	0.89
2049	Premonsoon	445	5.20	29.74	0.92
2050	Premonsoon	446	5.44	29.78	0.96
2051	Premonsoon	448	5.91	29.82	1.00
2052	Premonsoon	449	6.15	29.85	1.03
2053	Premonsoon	450	6.38	29.89	1.07
2054	Premonsoon	451	6.62	29.93	1.11
2055	Premonsoon	453	7.09	29.96	1.14
2056	Premonsoon	454	7.33	30.00	1.18
2057	Premonsoon	455	7.57	30.04	1.22
2058	Premonsoon	457	8.04	30.07	1.25
2059	Premonsoon	458	8.27	30.11	1.29

Year	Season	Rainfall (mm)	% Change of Rainfall with respect to 2024	Temperature (Degree Celsius)	Change of Temperature (Degree Celsius) with respect to 2024
2060	Premonsoon	459	8.51	30.15	1.33
2061	Premonsoon	461	8.98	30.19	1.37
2062	Premonsoon	462	9.22	30.22	1.40
2063	Premonsoon	464	9.69	30.26	1.44
2064	Premonsoon	465	9.93	30.30	1.48
2065	Premonsoon	467	10.40	30.33	1.51
2066	Premonsoon	468	10.64	30.37	1.55
2067	Premonsoon	470	11.11	30.41	1.59
2068	Premonsoon	472	11.58	30.44	1.62
2069	Premonsoon	473	11.82	30.48	1.66
2070	Premonsoon	475	12.29	30.52	1.70
2071	Premonsoon	477	12.77	30.55	1.73
2072	Premonsoon	478	13.00	30.59	1.77
2073	Premonsoon	480	13.48	30.63	1.81
2074	Premonsoon	482	13.95	30.67	1.85
2025	Monsoon	2126	0.33	31.34	0.04
2026	Monsoon	2133	0.66	31.38	0.08
2027	Monsoon	2140	0.99	31.41	0.11
2028	Monsoon	2147	1.32	31.45	0.15
2029	Monsoon	2154	1.65	31.48	0.18
2030	Monsoon	2160	1.93	31.52	0.22
2031	Monsoon	2166	2.22	31.56	0.26
2032	Monsoon	2173	2.55	31.59	0.29
2033	Monsoon	2179	2.83	31.63	0.33
2034	Monsoon	2185	3.11	31.66	0.36
2035	Monsoon	2191	3.40	31.70	0.40
2036	Monsoon	2197	3.68	31.74	0.44
2037	Monsoon	2203	3.96	31.77	0.47
2038	Monsoon	2208	4.20	31.81	0.51
2039	Monsoon	2214	4.48	31.85	0.55
2040	Monsoon	2219	4.72	31.88	0.58
2041	Monsoon	2225	5.00	31.92	0.62
2042	Monsoon	2230	5.24	31.95	0.65
2043	Monsoon	2235	5.47	31.99	0.69
2044	Monsoon	2240	5.71	32.03	0.73
2045	Monsoon	2245	5.95	32.06	0.76
2046	Monsoon	2250	6.18	32.10	0.80

Year	Season	Rainfall (mm)	% Change of Rainfall with respect to 2024	Temperature (Degree Celsius)	Change of Temperature (Degree Celsius) with respect to 2024
2047	Monsoon	2255	6.42	32.13	0.83
2048	Monsoon	2259	6.61	32.17	0.87
2049	Monsoon	2264	6.84	32.21	0.91
2050	Monsoon	2268	7.03	32.24	0.94
2051	Monsoon	2272	7.22	32.28	0.98
2052	Monsoon	2277	7.46	32.32	1.02
2053	Monsoon	2281	7.65	32.35	1.05
2054	Monsoon	2285	7.83	32.39	1.09
2055	Monsoon	2289	8.02	32.42	1.12
2056	Monsoon	2292	8.16	32.46	1.16
2057	Monsoon	2296	8.35	32.50	1.20
2058	Monsoon	2300	8.54	32.53	1.23
2059	Monsoon	2303	8.68	32.57	1.27
2060	Monsoon	2306	8.82	32.60	1.30
2061	Monsoon	2310	9.01	32.64	1.34
2062	Monsoon	2313	9.16	32.68	1.38
2063	Monsoon	2316	9.30	32.71	1.41
2064	Monsoon	2319	9.44	32.75	1.45
2065	Monsoon	2321	9.53	32.78	1.48
2066	Monsoon	2324	9.67	32.82	1.52
2067	Monsoon	2327	9.82	32.86	1.56
2068	Monsoon	2329	9.91	32.89	1.59
2069	Monsoon	2332	10.05	32.93	1.63
2070	Monsoon	2334	10.15	32.97	1.67
2071	Monsoon	2336	10.24	33.00	1.70
2072	Monsoon	2338	10.34	33.04	1.74
2073	Monsoon	2340	10.43	33.07	1.77
2074	Monsoon	2342	10.52	33.11	1.81
2025	Postmonsoon	157	0.00	26.81	0.04
2026	Postmonsoon	156	-0.64	26.84	0.07
2027	Postmonsoon	156	-0.64	26.87	0.10
2028	Postmonsoon	156	-0.64	26.90	0.13
2029	Postmonsoon	155	-1.27	26.94	0.17
2030	Postmonsoon	155	-1.27	26.97	0.20
2031	Postmonsoon	155	-1.27	27.00	0.23
2032	Postmonsoon	154	-1.91	27.03	0.26
2033	Postmonsoon	154	-1.91	27.07	0.30

Year	Season	Rainfall (mm)	% Change of Rainfall with respect to 2024	Temperature (Degree Celsius)	Change of Temperature (Degree Celsius) with respect to 2024
2034	Postmonsoon	154	-1.91	27.10	0.33
2035	Postmonsoon	153	-2.55	27.13	0.36
2036	Postmonsoon	153	-2.55	27.16	0.39
2037	Postmonsoon	153	-2.55	27.20	0.43
2038	Postmonsoon	152	-3.18	27.23	0.46
2039	Postmonsoon	152	-3.18	27.26	0.49
2040	Postmonsoon	152	-3.18	27.30	0.53
2041	Postmonsoon	151	-3.82	27.33	0.56
2042	Postmonsoon	151	-3.82	27.36	0.59
2043	Postmonsoon	151	-3.82	27.39	0.62
2044	Postmonsoon	150	-4.46	27.43	0.66
2045	Postmonsoon	150	-4.46	27.46	0.69
2046	Postmonsoon	150	-4.46	27.49	0.72
2047	Postmonsoon	149	-5.10	27.52	0.75
2048	Postmonsoon	149	-5.10	27.56	0.79
2049	Postmonsoon	149	-5.10	27.59	0.82
2050	Postmonsoon	148	-5.73	27.62	0.85
2051	Postmonsoon	148	-5.73	27.65	0.88
2052	Postmonsoon	148	-5.73	27.69	0.92
2053	Postmonsoon	147	-6.37	27.72	0.95
2054	Postmonsoon	147	-6.37	27.75	0.98
2055	Postmonsoon	147	-6.37	27.79	1.02
2056	Postmonsoon	146	-7.01	27.82	1.05
2057	Postmonsoon	146	-7.01	27.85	1.08
2058	Postmonsoon	146	-7.01	27.88	1.11
2059	Postmonsoon	146	-7.01	27.92	1.15
2060	Postmonsoon	145	-7.64	27.95	1.18
2061	Postmonsoon	145	-7.64	27.98	1.21
2062	Postmonsoon	145	-7.64	28.01	1.24
2063	Postmonsoon	144	-8.28	28.05	1.28
2064	Postmonsoon	144	-8.28	28.08	1.31
2065	Postmonsoon	144	-8.28	28.11	1.34
2066	Postmonsoon	143	-8.92	28.14	1.37
2067	Postmonsoon	143	-8.92	28.18	1.41
2068	Postmonsoon	143	-8.92	28.21	1.44
2069	Postmonsoon	142	-9.55	28.24	1.47
2070	Postmonsoon	142	-9.55	28.28	1.51

Year	Season	Rainfall (mm)	% Change of Rainfall with respect to 2024	Temperature (Degree Celsius)	Change of Temperature (Degree Celsius) with respect to 2024
2071	Postmonsoon	142	-9.55	28.31	1.54
2072	Postmonsoon	141	-10.19	28.34	1.57
2073	Postmonsoon	141	-10.19	28.37	1.60
2074	Postmonsoon	141	-10.19	28.41	1.64
2025	Winter	221	0.00	22.64	0.07
2026	Winter	222	0.45	22.71	0.14
2027	Winter	222	0.45	22.79	0.22
2028	Winter	222	0.45	22.86	0.29
2029	Winter	222	0.45	22.93	0.36
2030	Winter	222	0.45	23.00	0.43
2031	Winter	222	0.45	23.07	0.50
2032	Winter	223	0.90	23.15	0.58
2033	Winter	223	0.90	23.22	0.65
2034	Winter	223	0.90	23.29	0.72
2035	Winter	223	0.90	23.36	0.79
2036	Winter	223	0.90	23.44	0.87
2037	Winter	223	0.90	23.51	0.94
2038	Winter	223	0.90	23.58	1.01
2039	Winter	224	1.36	23.65	1.08
2040	Winter	224	1.36	23.72	1.15
2041	Winter	224	1.36	23.80	1.23
2042	Winter	224	1.36	23.87	1.30
2043	Winter	224	1.36	23.94	1.37
2044	Winter	224	1.36	24.01	1.44
2045	Winter	224	1.36	24.09	1.52
2046	Winter	225	1.81	24.16	1.59
2047	Winter	225	1.81	24.23	1.66
2048	Winter	225	1.81	24.30	1.73
2049	Winter	225	1.81	24.37	1.80
2050	Winter	225	1.81	24.45	1.88
2051	Winter	225	1.81	24.52	1.95
2052	Winter	225	1.81	24.59	2.02
2053	Winter	226	2.26	24.66	2.09
2054	Winter	226	2.26	24.74	2.17
2055	Winter	226	2.26	24.81	2.24
2056	Winter	226	2.26	24.88	2.31
2057	Winter	226	2.26	24.95	2.38

Year	Season	Rainfall (mm)	% Change of Rainfall with respect to 2024	Temperature (Degree Celsius)	Change of Temperature (Degree Celsius) with respect to 2024
2058	Winter	226	2.26	25.03	2.46
2059	Winter	227	2.71	25.10	2.53
2060	Winter	227	2.71	25.17	2.60
2061	Winter	227	2.71	25.24	2.67
2062	Winter	227	2.71	25.31	2.74
2063	Winter	227	2.71	25.39	2.82
2064	Winter	227	2.71	25.46	2.89
2065	Winter	227	2.71	25.53	2.96
2066	Winter	228	3.17	25.60	3.03
2067	Winter	228	3.17	25.68	3.11
2068	Winter	228	3.17	25.75	3.18
2069	Winter	228	3.17	25.82	3.25
2070	Winter	228	3.17	25.89	3.32
2071	Winter	228	3.17	25.96	3.39
2072	Winter	228	3.17	26.04	3.47
2073	Winter	229	3.62	26.11	3.54
2074	Winter	229	3.62	26.18	3.61

5.3 Comparative Analysis

From the Table 5.4, for 2050, as a sample analysis, it is shown that the increase of percentage of seasonal rainfall change with respect to 2024 is maximum for monsoon season and the change of seasonal average temperature in degree Celsius with respect to 2024 is maximum for winter season. The graphs showing the percentage change of rainfall between 2024 and 2050 and change of temperature in degree Celsius between 2024 and 2050 are presented in Figure 5.1 and Figure 5.2 respectively.

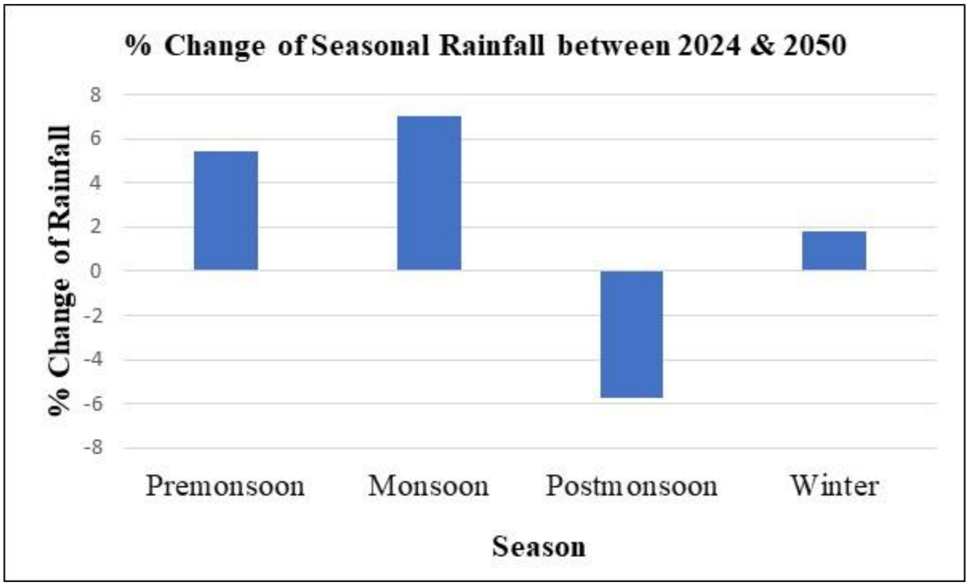

Figure 5.1: Percentage change of seasonal rainfall between 2024 & 2050

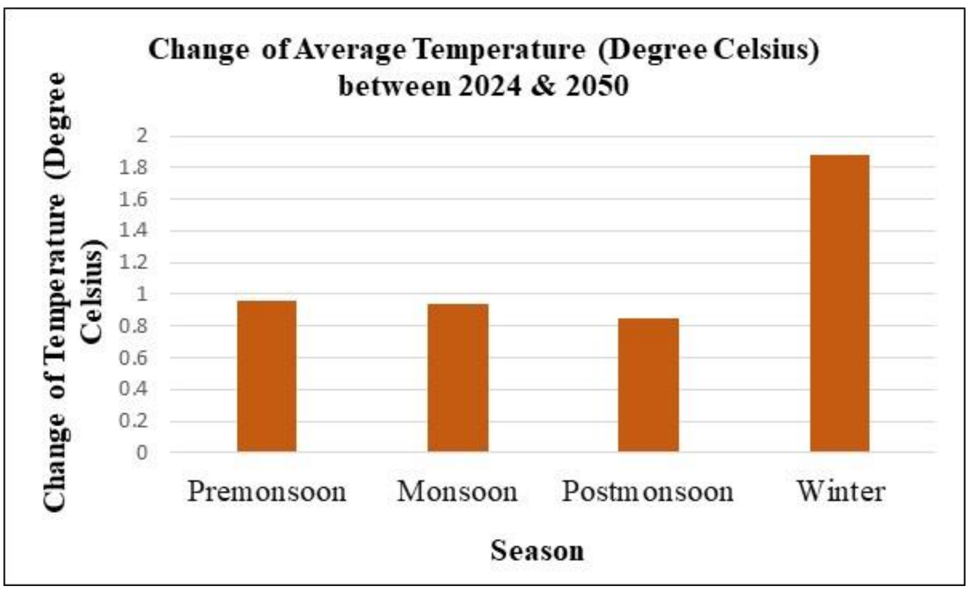

Figure 5.2: Change of seasonal temperature between 2024 & 2050

5.4 Summary

The study presents a comprehensive approach to predicting future seasonal rainfall and temperature in Bangladesh using machine learning models. Historical climate data from 188 locations across Bangladesh is processed to create two unified time series datasets: one for daily rainfall and another for daily average temperature, spanning from 1951 to 2014. These datasets are categorized into four seasons: Premonsoon, monsoon, Postmonsoon, and winter. For each season, the yearly sum of rainfall and the yearly average temperature are calculated, resulting in eight distinct datasets. Six machine learning models - linear regression, polynomial regression, decision tree, random forest, adaboost and gradient boosting machines - are developed for each dataset, leading to a total of 32 models.

The performance of each model was evaluated based on accuracy, and the best-performing model for each dataset is selected. This selection process is crucial to ensure that the most reliable and accurate predictions can be made for future climate scenarios. The selected models are then used to forecast seasonal rainfall and temperature for the next 50 years, from 2025 to 2074. This predictive analysis aims to provide insights into future climate trends, highlighting potential changes in seasonal patterns that can have significant implications for agriculture, water resources, and overall climate resilience in Bangladesh.

It can be said from the predicted future seasonal rainfall and average temperature for next 50 years (2025-2074) that rainfall will tend to increase in premonsoon, monsoon and winter season while the rainfall will be in decreasing trend in postmonsoon season. The temperature will tend to increase in all seasons such as premonsoon, monsoon, postmonsoon and winter season. The increased temperature will not be good for Bangladesh because it will warm the overall environment and cause various types of environmental problems. The increase of rainfall will also be threat to the different sectors of Bangladesh such as agriculture and water environment.

CHAPTER 6

Impact on Society, Environment and Sustainability

6.1 Impact on Life

The impact of accurate climate predictions derived from this thesis extends across various facets of life in Bangladesh. By forecasting future changes in seasonal rainfall and temperature with enhanced precision, stakeholders in agriculture can better plan crop cycles, irrigation strategies, and pest control measures. This proactive approach mitigates risks associated with climate variability, ensuring food security and stable agricultural yields. Moreover, water resource managers benefit from insights into expected precipitation patterns, enabling efficient allocation and management of water supplies for domestic, industrial, and agricultural use. This, in turn, supports sustainable development and resilience against droughts or floods, critical in a country prone to monsoonal variability.

Beyond agriculture and water management, the thesis findings inform public health strategies by predicting temperature-related health risks and outbreaks of climate-sensitive diseases. Authorities can implement timely interventions and public awareness campaigns to minimize health impacts during extreme weather events. Additionally, industries reliant on climate conditions, such as energy and tourism, can strategically plan operations and investments based on anticipated weather patterns. Overall, the thesis contributes to enhancing societal resilience to climate change in Bangladesh by empowering decision-makers with actionable insights derived from robust scientific predictions.

6.2 Impact on Society & Environment

Impact on Society:

The thesis on predicting future changes in seasonal rainfall and temperature in Bangladesh using machine learning models holds significant societal implications. As climate change continues to affect weather patterns globally, Bangladesh, with its high vulnerability to climatic fluctuations, stands to benefit immensely from accurate and reliable climate forecasts. By providing detailed predictions for the next 50 years, this research empowers policymakers and planners to devise informed strategies that mitigate the adverse impacts of climate variability on various sectors such as agriculture, water resources, public health, and infrastructure.

In the agricultural sector, which forms the backbone of Bangladesh's economy and employs a significant portion of its population, precise climate predictions are invaluable. The research highlights potential changes in rainfall and temperature patterns, enabling farmers to make informed decisions about crop selection, planting schedules, and irrigation practices. This foresight can reduce crop failure risks, enhance food security, and stabilize the livelihoods of millions of farmers. Moreover, the ability to anticipate extreme weather events, such as floods and droughts, allows for the implementation of proactive measures to protect crops and livestock, thereby minimizing economic losses.

Water resource management is another critical area that stands to benefit from the insights provided by this thesis. Bangladesh is prone to both water scarcity and excess, making effective water management crucial for sustaining its population and industries. The predictions of seasonal rainfall changes enable better planning for water storage, distribution, and flood control systems. This ensures a stable water supply for domestic, agricultural, and industrial use, and helps prevent the detrimental effects of water-related disasters. Improved water management also supports public health by ensuring access to clean water and reducing the incidence of waterborne diseases.

Public health is directly impacted by changes in climate, with temperature and rainfall influencing the spread of diseases, heat-related illnesses, and overall living conditions. The thesis' findings on temperature increases and altered rainfall patterns can guide public health initiatives to address these challenges. For instance, anticipating higher temperatures and increased rainfall during certain seasons can lead to enhanced preparedness for heatwaves and vector-borne diseases such as malaria and dengue. Public awareness campaigns, healthcare infrastructure improvements, and targeted interventions can be planned more effectively, reducing health risks and improving community resilience.

Finally, the thesis fosters a broader understanding of climate change's implications, promoting awareness and engagement within society. By disseminating knowledge about potential future climate scenarios, the research encourages communities, businesses, and individuals to adopt sustainable practices and support environmental conservation efforts. This collective awareness and proactive approach contribute to building a resilient society capable of adapting to and mitigating the impacts of climate change. Overall, the thesis not only advances scientific understanding but also provides practical tools and insights that are essential for safeguarding Bangladesh's socio-economic stability and well-being in the face of a changing climate.

Impact on Environment

The thesis on predicting future changes in seasonal rainfall and temperature in Bangladesh using machine learning models significantly impacts environmental conservation and management. By providing long-term climate forecasts, the research helps identify potential environmental challenges and opportunities. Accurate predictions of rainfall and temperature patterns enable the development of effective strategies to manage and protect natural resources, contributing to the preservation of ecosystems and biodiversity in Bangladesh.

One of the primary environmental impacts of this thesis is its contribution to water resource management. Bangladesh is highly vulnerable to both flooding and drought, and the research offers crucial insights into future rainfall patterns. This information can guide the construction and maintenance of water infrastructure, such as reservoirs, dams, and drainage systems, to prevent water-related disasters. Effective water management ensures that ecosystems reliant on specific water conditions remain stable, preserving habitats for aquatic and terrestrial species.

The predicted increase in temperature across all seasons, as highlighted in the thesis, poses significant challenges to Bangladesh's flora and fauna. Higher temperatures can lead to habitat loss, altered species distribution, and increased vulnerability to pests and diseases. By anticipating these changes, conservation efforts can be directed towards protecting vulnerable species and ecosystems. For instance, establishing wildlife corridors and protected areas can help species migrate to more suitable habitats, thereby maintaining biodiversity.

Additionally, the thesis addresses the potential for increased rainfall in the Premonsoon, Monsoon, and Winter seasons, which can lead to soil erosion, landslides, and sedimentation in water bodies. These processes can degrade land and aquatic habitats, affecting both plant and animal life. By forecasting these changes, land management practices can be adjusted to minimize erosion and protect soil health. Techniques such as reforestation, terracing, and sustainable agriculture can be promoted to enhance soil stability and fertility.

The research also emphasizes the importance of sustainable agricultural practices in the face of changing climate conditions. With accurate predictions of future climate patterns, farmers can adopt practices that reduce environmental impact, such as crop rotation, conservation tillage, and the use of drought-resistant crop varieties. These practices not only enhance agricultural productivity but also contribute to the overall health of the environment by reducing soil degradation, conserving water, and maintaining biodiversity. Finally, the thesis promotes environmental awareness and education by highlighting the critical need for proactive climate adaptation and mitigation strategies. It encourages policymakers, environmentalists, and the general public to engage in sustainable practices and support conservation initiatives. This collective effort is essential for building resilience against climate change, protecting natural ecosystems, and ensuring the long-term sustainability of Bangladesh's environmental resources. The research's findings serve as a valuable resource for guiding environmental policies and actions that aim to mitigate the adverse effects of climate change and foster a more sustainable future.

6.3 Ethical Aspects

The ethical aspects of this thesis primarily revolve around the responsible use of data and the implications of its findings on society and the environment. Ensuring the accuracy and reliability of climate predictions is crucial, as these forecasts directly influence policy decisions and adaptation strategies that affect millions of people. The use of historical climate data from 188 locations across Bangladesh necessitates stringent data handling protocols to maintain integrity, privacy, and accuracy. Any misrepresentation or error in the data could lead to misguided policies, potentially exacerbating the adverse impacts of climate change on vulnerable populations and ecosystems.

Additionally, the thesis underscores the ethical responsibility to communicate findings transparently and effectively to all stakeholders, including policymakers, farmers, and the general public. Clear communication of the risks and uncertainties associated with climate predictions is essential to foster informed decision-making and promote equitable adaptation strategies. It is also imperative to consider the ethical implications of proposed adaptation measures, ensuring they do not disproportionately disadvantage any community or ecosystem. By addressing these ethical considerations, the research not only contributes to scientific knowledge but also supports the development of fair and sustainable climate resilience strategies.

6.4 Sustainability Plan

The sustainability plan of this thesis focuses on ensuring that the research findings and methodologies have a lasting impact on both the scientific community and practical applications in Bangladesh. To begin with, the data collection and processing methods are designed to be replicable and scalable. By meticulously documenting the steps taken to create unified time series datasets for rainfall and temperature, future researchers can apply similar methodologies to other regions or datasets. This contributes to a growing body of knowledge on climate prediction and allows for continuous improvement and refinement of predictive models as new data becomes available.

A critical component of the sustainability plan is the engagement and collaboration with local stakeholders, including policymakers, farmers, environmentalists, and community leaders. By involving these stakeholders in the research process and dissemination of findings, the thesis ensures that the insights gained are directly translated into actionable strategies. Training sessions, workshops, and public seminars can be organized to educate stakeholders on using the predictive models and interpreting their results. This empowers local communities to make informed decisions regarding agriculture, water resource management, and disaster preparedness, thereby enhancing the resilience of these communities to climate change.

Finally, the sustainability plan emphasizes the need for continuous monitoring and evaluation of the implemented adaptation strategies. Establishing a framework for periodic review of climate predictions and their real-world impacts ensures that the models remain relevant and accurate over time. Feedback mechanisms can be set up to gather data on the effectiveness of the adaptation measures, allowing for iterative improvements. By fostering an environment of ongoing research, stakeholder engagement, and adaptive management, the thesis aims to create a sustainable approach to climate resilience that can evolve in response to new challenges and opportunities presented by climate change.

6.5 Summary

The thesis on predicting future changes in seasonal rainfall and temperature in Bangladesh through advanced machine learning models has profound impacts across multiple dimensions. It significantly enhances agricultural planning by enabling better crop cycle management, irrigation strategies, and pest control measures, thus ensuring food security and stable yields amidst climate variability. Moreover, it empowers water resource managers with insights into precipitation patterns, facilitating efficient allocation and management of water supplies for domestic, industrial, and agricultural use, crucial for sustainable development and resilience against floods and droughts. The findings also inform public health strategies to mitigate temperature-related health risks and outbreaks of climate-sensitive diseases, while aiding industries in adapting operations to anticipated weather patterns. Societally, the thesis supports informed policymaking across sectors, promoting resilience to climate change through actionable insights and fostering environmental conservation efforts. Ethically, the research emphasizes responsible data use and transparent communication to ensure equitable adaptation strategies, while its sustainability plan ensures ongoing relevance and applicability of its methodologies and findings through stakeholder engagement and continuous evaluation.

CHAPTER 7
Conclusion and Future Work

7.1 Conclusions

The conclusions of this study highlight the critical importance and effectiveness of using machine learning models to predict future seasonal rainfall and temperature changes in Bangladesh. By meticulously processing historical climate data from 1951 to 2014 and employing six distinct machine learning models - linear regression, polynomial regression, decision tree, random forest, adaboost and gradient boosting machines - the research successfully identifies the most accurate models for forecasting climate variables. The selected models reveal significant trends: an increase in rainfall is anticipated during the Premonsoon, Monsoon and winter seasons, while a decrease is expected in the Postmonsoon season. Additionally, temperature is projected to rise across all seasons, underscoring the pervasive impact of climate change on the region. These predictions are crucial for informing policy decisions and adaptive strategies in various sectors, including agriculture, water resource management, and public health, thereby enhancing Bangladesh's resilience to climate-related challenges.

The study's findings emphasize the need for immediate and informed action to mitigate the adverse effects of climate change. The anticipated increase in temperature and changes in rainfall patterns necessitate the adoption of adaptive agricultural practices to ensure food security and sustainable farming. Effective water management strategies must be implemented to prevent water scarcity and flooding, protecting both human populations and ecosystems. Public health initiatives should be strengthened to address the rising risks associated with heatwaves and vector-borne diseases. The research also highlights the ethical responsibility of accurately communicating these predictions to stakeholders and involving them in the development of practical adaptation measures. By fostering a collaborative approach and ensuring continuous monitoring and evaluation of implemented strategies, the study aims to create a sustainable framework for climate resilience. Ultimately, this research provides a comprehensive foundation for future studies and policy-making, contributing to a more prepared and adaptable society in Bangladesh, capable of facing the evolving challenges of climate change.

7.2 Future Suggested Works

The findings of this study open numerous avenues for further research to enhance the understanding and prediction of climate change impacts on Bangladesh. One primary implication is the need for incorporating more complex and diverse machine learning models, such as neural networks and deep learning algorithms, to potentially improve prediction accuracy and capture nonlinear relationships in the climate data. Future studies could also benefit from integrating additional climate variables, such as humidity, wind speed, and solar radiation, which could provide a more comprehensive understanding of the climatic shifts. By expanding the scope of variables and employing advanced machine learning techniques, researchers can develop more robust models that offer greater precision and reliability in forecasting future climate scenarios.

Additionally, further research should focus on downscaling global climate model projections to a finer regional and local scale to enhance the relevance and applicability of predictions. This involves using regional climate models (RCMs) in conjunction with machine learning approaches to refine predictions and provide more detailed insights into local climatic changes. Such high-resolution climate projections are crucial for developing localized adaptation strategies that are tailored to the specific needs and vulnerabilities of different regions within Bangladesh. Moreover, future studies could explore the socio-economic impacts of predicted climate changes, examining how various sectors such as agriculture, health, and infrastructure might be affected and identifying potential measures to mitigate these impacts.

Another significant area for future research is the assessment of the effectiveness of implemented adaptation strategies based on the predictions provided by the current study. Longitudinal studies that monitor the outcomes of adaptive measures over time can provide valuable feedback on their efficacy and guide further refinements. Additionally, collaborative research involving stakeholders from government, academia, and local communities can foster the co-creation of knowledge and ensure that adaptation strategies are both scientifically sound and practically feasible. By continually evaluating and adjusting strategies in response to new data and insights, future research can help build a more resilient and adaptive society in Bangladesh, capable of effectively responding to the evolving challenges posed by climate change.

7.3 Conflict of Interests:

In conducting research on predicting future changes in seasonal rainfall and temperature in Bangladesh using machine learning models, several potential conflicts of interest must be carefully considered and managed. Firstly, there could be conflicts related to data sources and funding. Depending on the source of funding or data access, there may be pressures or expectations that could influence the research outcomes or interpretations. It is crucial to maintain independence and integrity in data selection and analysis to avoid biases that could skew results.

Secondly, conflicts of interest could arise concerning the interpretation and dissemination of findings. Stakeholders, including policymakers, industries, and advocacy groups, may have diverse interests in how climate predictions are portrayed and utilized. Balancing scientific rigor with the practical implications of the research requires transparency in reporting uncertainties and limitations, ensuring that conclusions are evidence-based rather than influenced by external pressures.

Moreover, conflicts may emerge regarding the application of research outcomes. For instance, stakeholders in agriculture or water management sectors might have differing perspectives on how predictive models should inform decision-making. It is essential to navigate these interests impartially, prioritizing the public good and sustainable development goals over sector-specific agendas.

Lastly, conflicts could surface concerning the broader societal impacts of the research. Anticipating and addressing potential unintended consequences of climate forecasts, such as socioeconomic disparities in adaptation capabilities or environmental impacts, requires ethical foresight and proactive engagement with affected communities.

Overall, managing conflicts of interest in this thesis involves maintaining scientific rigor, transparency, and a commitment to ethical conduct throughout the research process. By navigating these potential conflicts with integrity and accountability, the study aims to contribute responsibly to understanding and addressing climate change impacts in Bangladesh.

REFERENCES

[1] Sashank, C., Raj, K. U., Amin, T. J., & Ayyasamy, S. (2023). Rainfall Prediction using Deep Learning and Machine LearningTechniques. Research Square (Research Square).

[2] Gujanatti Rudrappa,Nataraj Vijapur,Rajesh Pattar,Ravi Rathod,Rashmi Kulkarni(2021). Machine Learning Models Applied for Rainfall Prediction. Volume 11.

[3] Mekanik, F., Imteaz, M.A., Gato-Trinidad, S., Elmahdi, A. (2013). "Multiple Regression and Artificial Neural Network for Long-term Rainfall Forecasting Using Large-Scale Climate Modes." Journal of Hydrology, Volume 503, October.

[4] Oswal, N. (2021, April 14). Predicting Rainfall using Machine Learning Techniques. TechRxiv.

[5] Azad, A., Kashi, H., Farzin, S., Singh, V. P., Kisi, O., Karami, H., Sanikhani, H. (2020). "Novel approaches for air temperature prediction: A comparison of four hybrid evolutionary fuzzy models." Meteorological Applications, 27, e1817.

[6] Kumar, V., Yadav, V. K., & Dubey, E. S. (2022). Rainfall Prediction using Machine Learning. International Journal for Research in Applied Science and Engineering Technology, 10(5),2494–2497.

[7] Oyana, T. J. (2020). Spatial Analysis with R: Statistics, Visualization, and Computational Methods. CRC Press.

[8] Liyew, C.M., Melese, H.A(2021) Machine learning techniques to predict daily rainfall amount. J Big Data 8, 153.

[9] Mr Sandeep, Ms Jahnavi, K Asst, Prof. (2020). Title: Rainfall Prediction Using Machine Learning Techniques and an Analysis of the Outcomes of These Techniques. International Journal of Engineering Applied Sciences and Technology, 2020 Vol. 4, Issue 9, ISSN No. 2455-2143, Pages 365-371

[10] Singla, P., Jain, H., & Kumar, Y. (2023). Rainfall prediction using AI and ML. International journal of creative research thoughts, Volume 11, issue 5.

[11] Alizadeh, M. J., Kavianpour, M. R., Kisi, O., Nourani, V. (2017). "A new approach for simulating and forecasting the rainfall-runoff process within the next two months." Journal of Hydrology, Vol. 548, May 2017.

[12] Byun, J., Jun, C., Kim, J., Cha, J., & Narimani, R. (2023). Deep Learning-based Rainfall Prediction using Cloud Image Analysis. IEEE Transactions on Geoscience and Remote Sensing, 1.

[13] Poornima, S., & Pushpalatha, M. (2021). Prediction of Rainfall Using Intensified LSTM Based Recurrent Neural Network with Weighted Linear Units. Atmosphere, 10(11), 668.

[14] Rahman, A., Abbas, S., Gollapalli, M., Ahmed, R., Aftab, S., Ahmad, M., Khan, M. A., et al. (2022). Rainfall Prediction System Using Machine Learning Fusion for Smart Cities. Sensors, 22(9), 3504. MDPI AG.

[15] Barrera-Animas, A. Y., Oyedele, L. O., Bilal, M, Akinosho, T. D., Delgado, J. M. D., & Akanbi, L. (2022). Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. Machine Learning with Applications, 7, 100204.

[16] Dutta, K., & P, G. (2020). Rainfall Prediction using Machine Learning and Neural Network. International Journal of Recent Technology and Engineering, 9(1), 1954–1961.

[17] Salcedo-Sanz, S., Deo, R. C., Carro-Calvo, L., Saavedra-Moreno, B. (2016). "Monthly prediction of air temperature in Australia and New Zealand with machine learning algorithms." Theoretical and Applied Climatology, Vol. 125, July 2016.

[18] Umay Shah, Sanjay Grag, (2018) Rainfall Prediction: Accuracy Enhancement Using Machine Learning and Forecasting Techniques, IEEE conference papers.

[19] Joseph., & Ratheesh, T. K. (2013). Rainfall prediction using data mining techniques. International Journal of Computer Applications, 83(8).

[20] Kumar, V., Yadav, V. K., & Dubey, E. S. (2022). Rainfall Prediction using Machine Learning. International Journal for Research in Applied Science and Engineering Technology, 10(5),2494–2497.

[21] Wang, Y., Bai, Y., Yang, L., Li, H. (2021). "Short time air temperature prediction using pattern approximate matching." Energy and Buildings, 244, 111036.

[22] Bindu, B. H. (2023). "Rainfall Prediction Using Machine Learning." International Journal of Research Publication and Reviews, Vol. 4, No. 12, pp. 4734-4738, December 2023.

[23] Vishwakarma, H. (2020). "Climate Change Analysis Using Machine Learning." International Journal of Science and Research (IJSR), ISSN: 2319-7064, Volume 9, Issue 8, August 2020.

[24] Azari, B., Hassan, K., Pierce, J., Ebrahimi, S. (2022). "Evaluation of Machine Learning Methods Application in Temperature Prediction." Computational Research Progress in Applied Science and Engineering, 8(1), 1-12. DOI: 10.52547/crpase.8.1.2747.

[26] Dataset obtained from https://zenodo.org/records/3873998#.YRIb9IgzZnJ

Appendix A

Course Outcomes, Complex Engineering Problems (EP) and Complex Engineering Activities (EA) Addressing

Title:
PREDICTING FUTURE CHANGE OF SEASONAL RAINFALL AND TEMPERATURE IN BANGLADESH USING MACHINE LEARNING MODELS

Student ID: 201-15-3363

CO Description for FYDP

CO	CO Descriptions	PO
	Phase -I	
CO1	Integrate recently gained and previously acquired knowledge to predict future seasonal rainfall and temperature for the Final Year Design Project (FYDP)	**PO1**
CO2	Analyze different aspects of the goals in designing a solution for this FYDP	**PO2**
CO3	Explore diverse problem domains through a literature review, delineate the issues, and establish the goals for the FYDP	**PO4**
CO4	Perform economic evaluation and cost estimation and employ suitable project management procedures throughout the development life cycle of the FYDP	**PO11**
	Phase -II	
CO5	Design and develop technical solutions and system components or processes that meet specified requirements, ensuring compliance with public health and safety standards, as well as considering cultural, socioeconomic, and environmental factors in this FYDP	**PO3**
CO6	Choose and apply appropriate methodologies, resources, and contemporary engineering and IT technologies to address complex engineering processes, encompassing prediction and modeling, while adhering to relevant constraints in this FYDP	**PO5**
CO7	Analyze societal, health, safety, legal, and cultural considerations, along with associated responsibilities, in the context of professional engineering practice and the resolution of this problem, employing logical reasoning guided by contextual understanding.	**PO6**
CO8	Comprehend and evaluate the enduring sustainability and impact of professional engineering endeavors in addressing intricate engineering challenges within social and environmental frameworks.	**PO7**
CO9	Implement ethical principles and adhere to professional standards and norms in this FYDP	**PO8**
CO10	Capable of operating proficiently both individually and as a team member or leader across diverse teams and interdisciplinary settings in this FYDP.	**PO9**

CO11	Proficiently communicate with the engineering community and broader society regarding complex engineering endeavors, including the ability to comprehend and generate comprehensive reports and design documentation, as well as provide and receive clear instructions throughout this FYDP.	**PO10**
CO12	Acknowledge the importance of self-directed and life-long learning within the evolving landscape of technology, and possess the readiness and capability to engage in lifelong learning endeavors.	**PO12**

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP), and Attainment of Complex Engineering Activities (EA)

Addressing CO (1 to 8), Knowledge Profile (K), Attainment of Complex Engineering Problems (EP):

SN	EP Definition	Attainment	CO	Justification (with Knowledge Profile)	References
1.	**EP1: Depth of Knowledge required**	Yes	CO1, CO2, CO3, CO5, CO6, CO7 and CO8	This project exemplifies **fundamental engineering principles (K3)** through the application of machine learning models to forecast future rainfall and temperature, as well as through the extraction and processing of data using Python programming. Additionally, the project showcases **specialist knowledge (K4)** by integrating an understanding of weather science, specifically focusing on rainfall and temperature, alongside foundational knowledge in statistics, prediction methodologies, and machine learning models.	**Page no:** [26-34] **Section:** [3.2] **Page no:** [1-9] **Section:** [1.2]
				The project incorporates **engineering practice and design (K5)** through the detailed Engineering Design Diagram of the entire Research System. It also addresses **engineering practice and technology (K6)** by utilizing various models such as the linear regression model, polynomial regression model, decision tree model, random forest model, adaboost and gradient boosting machines.	**Page no:** [26-34] **Section:** [3.2] **Page no:** [26-34, 39-52] **Section:** [3.2, 4.2]

				This project adheres to **K8 (Research Literature)** by integrating insights from recent studies, thereby enhancing the knowledge of predicting future seasonal rainfall and temperature with machine learning models. It demonstrates a thorough understanding of current methodologies.	**Page no:** [14-23] **Section:** [2.2, 2.3]
2.	**EP2: Range of Conflicting Requirements**	Yes	CO2, and CO7	This project addresses **EP-2** by identifying the challenges in predicting future seasonal rainfall and temperature, including the shortcomings of traditional methods and the intricacies of developing machine learning models. Through comparative analysis, it tackles issues related to understanding spatial distributions and provides insights for improving diagnostic methodologies.	**Page no:** [23-25] **Section:** [2.4]
3.	**EP3: Depth of analysis required**	Yes	CO2, and CO6	This project addresses **EP-3** by thoroughly comparing experimental results, emphasizing the predicted values of future seasonal rainfall and temperature, and analyzing the changes in these parameters relative to the current situation through detailed examination of the predicted data.	**Page no:** [55-64] **Section:** [5.2, 5.3]
4.	**EP4: Familiarity of Issues**	Yes	CO8	This project's interdisciplinary approach reaches beyond traditional boundaries in computer science and engineering, influencing fields such as weather science, data science, and statistics, thereby illustrating **EP-4**.	**Page no:** [14-20, 23-25] **Section:** [2.2, 2.4]
5.	**EP5: Extends of application codes**	No	CO5	N/A	N/A

6.	**EP6: Extends of stakeholders involved and conflicting**	No	CO8	N/A	N/A
7.	**EP7: Interdependence**	Yes	CO5	This project adopts a holistic approach that integrates diverse elements spanning data collection, statistical analysis, and methodological innovations to tackle complex challenges in forecasting future seasonal rainfall and temperature. This comprehensive strategy ensures the fulfillment of **EP-7**.	**Page no:** [26-34] **Section:** [3.2]

Addressing CO11 with Complex Engineering Activities (EA) [Some or all of the following]:

SN	EA Definition	Attainment	CO	Justification	References
1.	**EA1: Range of resources**	Yes	CO11	Our project leverages a robust array of resources, encompassing state-of-the-art high-performance computing infrastructure, powerful hardware and software setups, a sophisticated programming environment centered around Jupyter notebooks, meticulous data access and management protocols, well-annotated datasets, and stringent ethical considerations. These elements collectively underpin our systematic research efforts aimed at advancing the prediction of future seasonal rainfall and temperature using machine learning models.	**Page no:** [34] **Section:** [3.3]

2.	**EA2: Level of interaction**	No		N/A	N/A
3.	**EA3: Innovation**	No		N/A	N/A
4.	**EA4: Consequences for society and the environment**	Yes		This project makes a significant societal contribution by enabling proactive measures to prepare for potential extreme weather scenarios in agriculture, water resources management, public health, and infrastructure planning. By accurately predicting future seasonal rainfall and temperature patterns, it empowers stakeholders to implement adaptive strategies effectively. Furthermore, the project upholds environmental sustainability goals through the efficient utilization of computational resources and adheres strictly to ethical guidelines in handling weather data.	**Page no:** [66-71] **Section:** [6.1, 6.2, 6.4]
5.	**EA-5: Familiarity**	Yes		This project advances the current research landscape by exploring innovative methods to predict future seasonal rainfall and temperature. It introduces novel concepts and conducts thorough comparative analyses, thereby providing fresh perspectives and contributing new knowledge to the field.	**Page no:** [14-23] **Section:** [2.2, 2.3]

Addressing CO (4, 9, 10, and 12):

SN	COs	Attainment	Justification	References
1	CO4	Yes	This project meets **CO4** by implementing robust project management strategies and stringent financial oversight. It emphasizes careful planning, efficient resource allocation, and accurate budget forecasting to maximize resources effectively throughout the entire research process.	**Page no:** [35-37] **Section:** [3.4]
2	CO9	Yes	The project upholds ethical standards by placing paramount importance on the ethical handling of weather data, ensuring informed consent, and maintaining transparent documentation of the research procedures. These efforts are geared towards responsibly disseminating knowledge and promoting societal welfare through the ethical application of advanced weather prediction technologies, aligning with **CO9**.	**Page no:** [69-70] **Section:** [6.3]
3	CO10	No	N/A	N/A
4	CO12	Yes	The project exemplifies a commitment to ongoing learning (**CO12**) and adaptation in a rapidly evolving technological environment. This dedication is evident in its extensive data collection, rigorous statistical analysis, meticulous methodology development, and detailed experimental results and analysis. By continuously refining techniques and staying abreast of contemporary challenges, the project demonstrates a proactive approach to innovation and improvement in the field.	**Page no:** [26-34, 55-64] **Section:** [3.2, 3.3, 5.2, 5.3]

Plagiarism Check_201-15-3363.pdf

ORIGINALITY REPORT

24%	21%	14%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ijrpr.com Internet Source	5%
2	Submitted to Middlesex University Student Paper	2%
3	Submitted to Daffodil International University Student Paper	2%
4	Sunil Kumar. "Chapter 21 Polynomial Regression", Springer Science and Business Media LLC, 2024 Publication	1%
5	www.programcreek.com Internet Source	1%
6	www.researchgate.net Internet Source	1%
7	criticallycoding.home.blog Internet Source	1%
8	github.com Internet Source	1%