Use of Machine Learning to Predict and Analyze the Climatic Trends: A Data-Driven Approach

BY

Shenjuti Paul Puja

0242310005103037

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

Supervised By

Ms. Naznin Sultana

Associate Professor

Department of CSE

Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JULY 2024

APPROVAL

This Project titled **"Use of Machine Learning to Predict and Analyze the Climatic Trends: A Data-Driven Approach"** Submitted by Shenjuti Paul Puja, ID No: 0242310005103037 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 M.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 06-07-2024.

BOARD OF EXAMINERS

Chairman

Dr. Sheak Rashed Haider Noori, PhD
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. Md. Zahid Hasan, PhD
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. Arif Mahmud, PhD
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

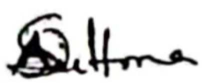
External Examiner

Dr. Md. Zulfiker Mahmud
Professor
Department of Computer Science and Engineering
Jagannath University

DECLARATION

I hereby declare that the project I worked on, entitled **"Use of Machine Learning to Predict and Analyze the Climatic Trends: A Data-Driven Approach"** Submitted to the Daffodil International University, is a record of original work done by me. As acknowledged in the text, except that the material has not been submitted, either in whole or in part, for a degree at this or any other university.

Supervised by:



Ms. Naznin Sultana
Associate Professor
Department of CSE
Daffodil International University

Submitted by:

Shejuti Paul Puja
0242310005103037
Department of CSE
Daffodil International University

ACKNOWLEDGMENT

First, I express my heartiest thanks and gratitude to almighty Allah, for His divine blessing makes it possible for me to complete this project successfully.

I am grateful to and wish to express my profound indebtedness to Supervisor **Ms. Naznin Sultana, Associate Professor**, Department of Computer Science and Engineering, Daffodil International University, Dhaka. My supervisor has deep knowledge and a deep interest in computer science to accomplish this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, and reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

I would like to express my heartfelt gratitude to Dr. Shek Rashed Haider Noori, Professor and Head, Department of CSE, Daffodil International University, for his kind help in finishing my project, as well as to other faculty members and the staff of the CSE department at Daffodil International University.

I would like to thank my entire coursemate at Daffodil International University, who took part in this discussion while completing the coursework.

Finally, I must acknowledge with due respect the constant support and patience of my parents.

ABSTRACT

Climate change is an alarming problem in the world, in recent years the extreme heat wave has been quite noticeable. For the last couple of years a temperate country like Bangladesh, is suffering from extreme heat waves. The temperature in many areas of Bangladesh rises to 40 degrees Celsius in the summer season. By investigating the data that have been collected for this study, tried to understand the weather trend and forest area trend of Bangladesh. The data was collected from various sources. NASA's POWER Project provides the temperature data, while the Humanitarian Data Exchange (HDX) website provides the rainfall data, with particular attention to Bangladeshi rainfall data. The forest area data was collected from World Bank Data on Forest Area and FAO Bangladesh Country Paper. 48 years of historical data have been used there. There are four types of data present in this dataset. These are Temperature(C), Date, Rain (millimeters), and Forest Area(percentage). 586 sets of data are available in this dataset. By investigating the dataset by exploratory data analysis it was found that there are positive high correlation between rainfall and temperature and a negative moderate correlation between forest area and year. Four machine-learning models were used there to do a comparative analysis between models. The models are gradient boosting, random forest, ridge regression, and lasso regression with an accuracy of 96.01 95.57, 94.07, and 93.89 respectively. The best-performing model was then integrated into a user-friendly web-based interface, allowing users to input specific years and months to receive temperature predictions. The interface also utilizes historical trends in rainfall and forest area to enhance prediction accuracy. This study fills in the current gaps in predictive modeling and provides everyone with a key resource for enhancing climate resilience strategies and making well-informed decisions.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction

In the last few years, global warming has sped up a lot, and the world`s climate decay is even more visible than before. Due to these alterations, Bangladesh which is found in the south-eastern South Asia region was affected so much than any other nation. The country is bordered by hilly wooded areas of Tripura, Mizoram(India), and Myanmar on the west-east sides; while to its north lies the majestic Himalayas; in contrast its southernmost point connects with the Bay of Bengal [1]. This area is vulnerable to climate change impact with frequent intense hot periods that are almost incessant now. In short, it is important to realize that in Bangladesh there is no silver bullet for the persistent hot weather conditions that we continue experiencing. Among the reasons behind this catastrophe are cutting down trees, melting arctic ice as well as discharging greenhouse gases into the atmosphere. Between 2010 and 2020 there were about 4.7 million hectares of forest loss each year across the globe. [2].However, more recent figures presented by the UN FAO show that deforestation is slightly higher – every year about 10 million hectares of forests are destroyed [3]. The Intergovernmental Panel on Climate Change (IPCC) asserts that the signification rise of the world mean temperature has been recorded at about 0. Changes the temperature by 85°C from 1880 to 2012 and 0. 78°C between the periods 1850-1900 and 2003-2012 [4]. These are seen to have brought deep consequent changes such as a record high rates of sea level increase that is rising at an average of 3. 2 mm per year. Also, the atmospheric concentration of carbon dioxide (CO2), one of the principal greenhouse gases, has risen over 25% in the last hundred years [5]. IPCC revealed in the Fifth Assessment Report that a 40% rise in the concentration of CO2 in the atmosphere is attributed to the burning of fossil and deforestation also known as land use change [4] [6].

Therefore, any global plan that aspires to control and store carbon, has to have forests in view since it's well known that they are sinks and reservoirs of CO2, which are also critical in setting global temperatures. However, the effects of climate change that are evident are likely to change the forest cover of the world shortly. this encompasses; species movement along the longitudinal and vertical axis of this planet as it tries to adjust to the pervading climatic conditions [7] [8]. This research paper has established the fact that the heat experienced beyond the comfort limit in different parts of the world has tripled due to deforestation and the unabated emission of greenhouse gases. These environmental changes make an urgent and combined effort to conserve and rehabilitate the forests paramount. Such valuable ecosystems should be preserved to reduce the impact of climate change, so important for the further existence of our planet. International collaboration and determination to prevent the degradation of the world's forests as well as controlling the emission of gases that cause global warming are critical to managing this problem.

This assessment is aimed at comparing the efficiency of different types of regression models in the context of forest areas' predictions based on the data on weather conditions. Other variables that are likely to affect the forest area and included in the dataset are temperature, month, year, and rainfall. Hence, using Ridge Regression, Lasso Regression, Random Forest, and Gradient Boosting models, it is our goal to establish which model allows for the most accurate predictions. Knowledge of each model's features will help to define the most suitable approach for capturing the interactions within data sets, therefore contributing to future improvements in forest, climate, and other resource management and preserves. The purpose of this analysis is to understand whether the linear or non-linear formats are more appropriate for environmental data of this kind.

1.2 Motivation

While there is always something in this world to cause alarm, global warming has always existed, but of late has had more evident impacts. Another dangerous phenomenon is the exacerbation of heat waves, about which scientists and researchers mostly indicate deforestation. Even though the notions concerning climate change and its causes have been discussed for thousands of years, these topics have become more pressing during the past several decades mainly because of the increase in the rate of global warming.

The motivation for this research stems from the urgent need to address the growing challenges posed by climate change and variability. Accurate temperature predictions are critical for effective environmental planning and management, yet traditional forecasting methods often fall short in capturing the intricate interactions between different climatic factors. By harnessing the power of machine learning and utilizing extensive historical datasets, we aim to improve the accuracy and reliability of temperature predictions. This research seeks to provide a robust tool that can help policymakers, environmentalists, and stakeholders make informed decisions to mitigate the adverse effects of climate change. Additionally, the development of a user-friendly interface ensures that these predictive capabilities are accessible to a broader audience, further empowering communities to anticipate and adapt to future climatic conditions.

1.3 The Rationale of the Study

Understanding temperature trends and rainfall variability helps identify localized effects of global warming, which can lead to extreme weather events impacting natural and human systems. Studying the relationship between rainfall and temperature in forests reveals ecosystem resilience to changing weather patterns, aiding conservation efforts. Accurate data informs environmental and climate policies, guiding decisions on forest preservation and climate change mitigation. Regional climate trend analysis helps develop targeted adaptation strategies, enhancing resilience. This study's analysis of weather and forest area data advances scientific understanding in ecology, climatology, and environmental science, providing benchmarks for future research and deepening knowledge of climate dynamics. This study aims to present a thorough examination of the links among rainfall, temperature, and forest area. Comprehending these dynamics will facilitate the progression of scientific understanding, resource allocation, policy development, the influence of climate variables, and climate change research. The results of this study are expected to have an impact on and provide direction for initiatives aimed at reducing climate change and advancing environmentally friendly behaviors.

1.4 Research Question

In particular, this study examines the relationship between average rainfall, forest area, and heat waves in connection to climate change. The results are addressed through the application of several machine learning models, and exploratory data analysis. The Questions are:

1. How can historical weather data and forest area records be effectively utilized to predict future temperatures?
2. Which machine learning models provide the highest accuracy in predicting future temperature trends based on historical data?
3. How can a user-friendly interface be developed to allow easy access to temperature predictions for future dates?

These inquiries aim to give a comprehensive understanding of the relationships between climatic variables and forest ecosystems, which will help to enhance strategies for environmental management and climate adaptation.

1.5 Expected Outcome

The outcome of this research is the development of a sophisticated machine learning model capable of accurately predicting future temperatures based on historical weather data and forest area records. By evaluating multiple models, including Ridge Regression, Lasso Regression, Random Forest, and Gradient Boosting, we identified the most accurate model for our dataset. This model was then integrated into a user-friendly web-based interface, allowing users to input specific years and months to receive temperature predictions. The interface also automatically incorporates historical trends in rainfall and forest area to enhance the accuracy of predictions. This tool not only demonstrates the potential of machine learning in climate prediction but also provides a practical solution for stakeholders involved in environmental planning and management, offering valuable insights for better decision-making and climate resilience strategies.

1.6 Research Objective

The primary goal of this research is to develop a robust and accurate predictive model for future temperatures using historical weather data and forest area records. This involves collecting and compiling comprehensive data from various online sources, encompassing variables such as year, month, rainfall, and forest area, with temperature as the target variable. We aim to evaluate and compare multiple machine learning models to identify the one that provides the highest prediction accuracy. Furthermore, we seek to create a user-friendly web-based interface that leverages the best-performing model to predict future temperatures based on user inputs for year and month. This interface is designed to make temperature predictions easily accessible and practical for environmental planning and management, thereby enhancing our ability to anticipate and respond to climate variability.

1.7 Problem Statement

The unpredictability of climate patterns poses significant challenges for environmental planning and management. Accurate temperature predictions are crucial for addressing these challenges, yet traditional methods often fail to capture the complex interactions between various climate factors. Despite the availability of extensive historical data on temperature, rainfall, and forest area, there is a lack of robust predictive models that leverage this data effectively. This research aims to address this gap by applying advanced machine learning techniques to historical weather and forest area data, providing a reliable tool for predicting future temperature trends.

1.8 Research Gap

Existing predictive models often fail to comprehensively incorporate the multi-dimensional relationships between temperature, rainfall, and forest area. Previous studies have typically focused on short-term predictions or limited geographic regions, lacking the generalizability and accuracy required for broader application. Moreover, there is a scarcity of user-friendly tools that enable stakeholders to easily access and utilize these predictions for practical decision-making. This research seeks to fill these gaps by developing a robust machine learning model trained on a long-term dataset and integrating it with an accessible web-based interface, thereby enhancing the accuracy and usability of temperature predictions and contributing to better climate resilience and environmental management.

1.9 Report Layout

The paper is divided into six chapters. Each chapter is examined from a variety of viewpoints and each chapter contains multiple components that are thoroughly explained. This report's content is given below:

Chapter 1

The following are some of the segments of this chapter - Discussion about the Introduction part, Discussion section about Motivations, Discuss the Rational Study of this article, Research Question, Expected Outcome.

Chapter 2

In this chapter, we discuss the Introduction, Related Works, Comparative analysis and Summary, Scope of the problem, and Challenges.

Chapter 3

In this chapter, we described the full working flow of our work. Including Some sections, Introduction, Research Subject and Instrumentations, Data Collection, Data Preprocessing, Statistical Analysis, Proposed Methodology, and Implementation Process.

Chapter 4

This chapter covers the Experiment and Results in a Discussion of this research, Introduction, Experimental Result & Analysis, Discussion.

Chapter 5

In This chapter, we discussed different social impacts in our society, Impact on Society, Impact on the Environment, Ethical Aspects, and Sustainability Plan

Chapter 6

This Section discussed the Summary of The Study, Conclusion, and Implications for Further Study.

CHAPTER 2
BACKGROUND

2.1 Introduction

The striking rise in heat wave frequency and intensity in recent years has created significant social, political, and environmental challenges. Although concerns about global warming date back thousands of years, recent data show that temperatures are rising more quickly than they have throughout history. There are currently severe heat waves in Bangladesh, a nation known for its generally moderate temperatures, which highlights this tendency. Bangladesh is a nation in South Asia that borders Myanmar to the southeast, India to the west, north, and east, and the Bay of Bengal to the south [31]. Together with other natural features like the Bay of Bengal, the Sundarbans mangrove forest, and the Chittagong Hill Tracts, this unique geographic location would normally result in a more moderate temperature. However, for the past five years, temperatures have continuously risen above 40 degrees Celsius, which is a marked divergence from historical patterns. In this study, the roles that deforestation, climate change, and altered rainfall patterns play in causing these unexpected and extreme heat waves are examined. The objective of the study is to comprehend how these factors interact to affect Bangladesh's heat wave frequency and intensity.

2.2 Related Work

Robert Mendelsohn's[12] research is a thorough examination of the effects of climate change on agriculture throughout Asia. The study uses a Ricardian model initially designed for China to interpolate the prospective consequences of climate change on agriculture across the continent. According to the research, a 1.5°C warming scenario may cut crop net earnings by 13%, resulting in a $93 billion loss per year. A more severe 3°C warming scenario is expected to quadruple these damages to $195 billion per year, resulting in a 28% decrease in net revenue. The report emphasizes the uneven distribution of repercussions, with countries such as India expected to suffer considerable losses, but other regions may face less impact due to varied meteorological and economic conditions. Although it does not completely balance the overall negative consequences, the inclusion of carbon fertilization effects somewhat mitigates these losses. In Asian agriculture, adaptive techniques are desperately needed to reduce the consequences of climate change and maintain food security. This research emphasizes this point.

More specifically, Malhi, Kaur, and Kaushik[13] review the climate changes that affect agriculture and undertake the discussion on measures that could minimize its impact. All the writers stress such points as increased temperature and CO_2 concentration influence on the soil and insects as well as on the plant and crop productivity. Increased photosynthesis productivity improves the stresses of temperature since when CO_2 rises, there is higher respiration and evapotranspiration. Other topics discussed include the impacts of climate change on access and utilization of health facilities among people with other ailments, pest build-up, and shifts in weed species. The measures that are discussed in this chapter include mitigation practices that are associated with climate-smart agriculture that focus on sustainability and food security. Resistance to drought, improved watering, and intensified utilization of the soil are recommended crops, watering, and the use of the soil in general. Such measures of technological, agronomic, and economic innovation are proposed by the authors to work in one symbiotic network to manage the negative impact of climate change on agriculture.

This is coupled with approaches for identifying regional climate responses to 1. 5°C and 2°C global warming targets are discussed with references to James et al[14]. Regional impacts are forecasted based on the various climate models and the proposed scenarios of changes in the climate. Based on the information of the analysis, the elevating global temperatures may cause some regional differences even if it is small in degree, which is something affecting climate extremes such as heatwaves, and heavy rainfalls. To make a better understanding of the impressions of climate change on a limited spatial scale and for the better determination of the alterations at the regional level, the authors underline the importance of high-resolution climate predictions. About increasing our understanding of climate effects at higher spatial resolution, the study calls for increased comparisons between models and the development of better resolution climate predictions at the regional level.

Even while mimicking the distribution of plants due to climate change, Oliveira et al. [15] have emphasized on the soil predictors. This makes a vast improvement in increasing the accuracy of ecological niche models through the addition of soil factors by applying machine learning algorithms. Brazil is the area of study, specifically the semi-arid area, and the study aims to predict what the impact of climate change will be on the vegetation of the area under varied climatic conditions. The findings illustrate this acute awareness of the need to assimilate soil components into climatic change computation to give better value estimates and predictions of the general environment's transformations. Therefore, the study concludes that in the case of climate change simple measurements of the soils are very important for the formulation of necessitating and effective conservation as well as land management plans.

When comparing the results of two generations of climate models – the CMIP3 and CMIP5, Lutz and his fellow authors[16] definitely focus on the impact on glaciers of the Central Asia. The two ensembles predict the increase in the temperature and amount of precipitation in the area to be quite large, though the models of CMIP5 are more steep, says the study. Aligned with the study's goal, the information underlines how significant climate models for adaptive planning are by presenting the effects of shifts in water resources in the region and glacier shrinkage. In their view, to control water availability and minimize the impact of glaciers' melting in Central Asia, long-term climatic forecasts suffixed.

The application of the current climate models to water resources is reviewed by Kundzewicz and Stakhiv[17]. The authors effectively enlighten the significant amount of risks that are still present despite explaining the strengths and weaknesses of the models. They argue that even when models provide valuable climate details, more research is required to enhance the models' accuracy and reliability, especially in the representation of the impacts of hydrology. As for advancing of water management and making policies in the framework of climate change, the report urges to enhance the development and validation of the models.

The multi-model implementation for increasing complexity in climate caliber models(MAGICC6) is brought about by Meinshausen and colleagues[18]. The objective of this condensed model is a scenario that requires the same predictions that are less processed. The calibration of the research is described in detail and the results indicate that MAGICC6 is a good representation of climate changes. Thus, the study shows that MAGICC6 can be effectively used for scenario analysis and policy making because it employs reduced compute resources than the complex models but develops adequate climate projections.

It elaborates on the HAPPI project that comparatively examines 1. Hence, this paper considers the impacts of an additional 0. 5°C of warming. The climate models and experimental setup used in this study are explained by Mitchell, et al .[19]. Conservative impacts on addressing the issue are the major findings of the research including resulting effects on species, sea levels, and extreme conditions. It also shows how emergent it is that world warm up not more than 1. Reduction of temperature by 5°C in a bid to mitigate impact of climate change. To protect the vulnerable communities and ecosystems CO, the study calls for more international cooperation to achieve more stringent climate goals.

Moss and associates provide updated climate change scenarios that have been developed for uses in analysis and assessment [20]. These cases provide a more comprehensive context that helps consider the possibility of climate risks since it takes into account economic and social factors and actions of governments. The writers describe the process and emphasize that it is more preferable than previous cases This is why, they underline how effective it is to use different sources and opinions. The cases made for advanced scenarios in this study come from the proposition that they are necessary in order to target efforts towards adaptation and mitigation and to offer more efficient solutions to the issues created by climate change.

The possible analogues that King and associates used the statistical extremes of Australian climate for analysing[21]. Implied below it, it is the 1. This means attaining a 5°C and 2°C target of global warming. At both warming levels an advanced climate model also suggests that heat waves, droughts and heavy rainfall events will become more frequent and/or intense in the future. According to the specific information published in the report, organisations from Australia are most threatened in terms of economy and environment, stating, one has to act without waiting in order to keep the worldwide increase in temperature below 1. 5°C. Since, advancement in climate change is making impacts more detectable, it is cardinal for ecosystems as well as communities to enhance the approaches of adaptation.

To understand the effects that are not encompassed by climate prevention measures, Sanderson and colleagues[22] have presented General Circulation Model (GCM) simulations for a new group of cases. Together with this, the increase in sea level, temperature, and climate variation in the second case is emphasized with reference to the results obtained with and without measures. By organizing the results one is able to conclude that aggressive measures in mitigating climate change could actually minimize the effects nearly to insignificant levels. The writers continue their dispute regarding the necessity of consistent efforts in the conception of reasonable climate policies as potential risks decreasing tools and the performance increasing factors for the populations.

Nelson and associates undertake an analysis of the literature to arrive at their meta-analysis on the financial costs of climate change on world agriculture[23]. It is necessary to clarify that because of the purposes of offering sufficient assessment of economic losses, the paper encompasses the results of several models and empirical researches. The consequence discovered include, negative impacts on crop yields and production which are mainly expressed in the developing states. In reference to the matter, crop types are presented as mirror images of a change strategy on the part of the writers of the crisis paper on agriculture in relation to climate change. As a result of climate change affecting food production, the study regards financing of agricultural research as an important area that must be expanded.

Wong et al. [24] discuss framework and approaches to the adaptation of dealing with climate change effects in the coastal regions. Thus, the authors of the study focus on the increased risk of coastal zones, erosion, storm surges and sea level rise. When it comes to the methods of adaptation, the writers describe both the soft and hard management solutions which include the EBA and the seawalls alike. This paper underscores how the engagement of stakeholders' participation and local knowledge is key towards formulating the ideal adaptation measures. Through the efforts of analysing the problems and deficiencies of coastal towns concerning their climate change adaptation processes, the authors provide a comprehensive strategy that ought to be adopted to facilitate the enhancement of the vulnerability of coastal towns.

Nakashima et al. [25] in their work on climate change adaptability blend scientific data and local knowledge. Also, the review insist on the fact that for recognizing the impacts of climate change and developing the mitigation strategies, the frameworks of the local and indigenous knowledge are crucial. Here, the writers provide some new case studies about how scientific research findings of climate change and indigenous knowledge can be combined in order to enhance the readiness of communities. As a part of the political strategy based on recognizing and including indigenous peoples' knowledge, the article calls for the expanded awareness of the importance of traditional knowledge in climate change adaptation policies and practices.

Bellard and associates[26] discuss on how change in climate influences the issues of biodiversity as well as ecological services. This paper contains the review of the study on changes in the distribution of species, the deterioration of habitats and alteration of the ecosystem processes. Besides, being very crucial and beneficial as pollinators, water purifiers, and carbon sinks among others, it is worrisome that the authors emphasis the importance of sustaining ecological commodity through the functioning of a wide variety of species. Depending on the urgency of the matter, the assessment advocates for the need to effect policy reforms in order to keep safeguarding the different species. Such are climate refugia, habitat restoration, and the integration of the conservation of the biological diversity in climate change adaptation and Response planning.

Lesnikowski and associates[27] used the evidence to look at policy approaches to climate change while looking at national examples and practices as well as emerging trends. This consideration makes the review focus on the process of developing and implementing national adaptation plans and strategies concentrating on successes and common issues. The authors identify sources of financing, stakeholders' engagement, and governance structures, among other critical factors that determine the extent of effective adaptation policies. As for the main developed assessment principles, the applicable policy focus is made on the existence of adaptable governance and frameworks for policy changes in accordance with new knowledge and new climate conditions.

Tactics and methods of urban climate change adaptation are described by Revi and associates.[28] Cities are highlighted to be having specific challenges that they face like heat waves, flooding and weak infrastructures. The writers discussed several adaptation approaches that include disaster risk reduction, improving urban design and using the concept of green infrastructure. The review explains why climate adaptation should be incorporated into urban policies governing growth and why the local people should also be involved in this cause. The authors recommend further research concerning the resilience of cities and the development of adaptation programs specific to the context.

In the event of food vulnerability, Vermeulen, Campbell, and Ingram[29] also analyze innovation strategies on climate change. Precise crop varieties that are resistant to climate change, sustainable land management and water productivity are among the techniques which the authors insist are necessary. Some of these are the technology, community-based techniques and policy interferences that can applied in food security. To demonstrate the ideas of minimizing the impact of climate change on agriculture and ensuring food security, the authors describe numerous successful experiences of other regions.

There is a discussion on how climate resilience should be connected with sustainable development by Schipper and Pelling[30], who argue that there must be a synchronization of the social and environmental goals. When discussing the approaches to increase the environmental and community's levels of resilience, the studied work highlights the importance of recourse to governance and participation processes. According to the authors, advocating climate resilience and sustainable development simultaneously is relevant as they are mutually beneficial. Therefore, in order to enhance overall climate change preparedness, the review recommends policies on social justice, environmental conservation, and fair distribution of resources.

2.3 Comparative Analysis and Summary

The research papers give a general insight into the influence of climate change in various fields. As is projected in the agricultural business, with demands for such crops and practices, that are resistant to the unfavorable effects of climate change. Food security is threatened in the current world; hence, there is a need time and again to seek to develop diverse adaptation strategies. Climate models and projections are very necessary for actual future effects and need enhancement in their accuracy along with localization. Thus, water resources are not immune to problems resulting from fluctuating precipitation and melting glaciers; therefore, strong management measures are needed.

The general abundance and distribution of species as well as ecosystem services which are the basis for human livelihoods are affected by habitat loss and changes in species distribution hence the issue of conservation. Concerning the living conditions in cities, heat waves, floods, and damaged infrastructure are critical, and better planning and disaster risk management are needed. These adaptation strategies have to overcome the institutional impediments and blend the indigenous knowledge with the scientific.

The problems in terms of their impacts are widespread and embrace virtually all spheres of life, from agriculture and water resources to biological diversity, urban structures, and policy systems. Some of the issues for development include enhancing the accuracy of specificity of climate models, designing climate robust farming technologies, rational use of water, conserving biological diversity, enhancing the ability of cities to adapt to the effects of climate change, climate proofing governance, and flexible policies. Sharing of ideas and working together with other people, companies, and government departments, as well as listening to them and coming up with new ideas, measures, and products is very important in enhancing resilience, especially in the face of climate change in the mentioned sectors.

2.4 Scope of the problem

Climate change results in different issues. It affects agriculture through reduced crop productivity and this in turn affects the economy negatively. Availability of water is impacted by rainfall and glaciers in the area hence leading to water shortages. Biodiversity also endures fragmentation of habitats and alterations to ecosystems, and these services include pollination and water purification. These disasters affect cities more, especially with heatwaves, floods or, degradation of infrastructural facilities, meaning that proper and efficient urbanism and risk management or reduction is important. Adaptation is said to be constrained by policies, and adoption problems such as lack of resources and flexibility, and there is a vital essence of engaging local and traditional knowledge.

2.5 Challenges

Several challenges need addressing. It is therefore important to enhance the accuracy of the current climate models particularly for the regional ones. Genetically modifying crops for climate change or using water more effectively are some of the principles of importance in agriculture. Proper management of water and its supply and an ability to forecast the fluctuation are essential. That is why it is essential to ensure the preservation of biological diversity with the help of efficient conservation methods. The study concludes that it is relevant for urban areas to address the two aspects of structural strength and climate change preparedness in their development. It also stressed the need to overcome governance barriers and to establish more open and liberalist policies for good climate change action. Lastly, it is possible to conclude that the inclusion of traditional knowledge and stakeholders in the frame of scientific methods in the case of adaptation is critical. In general, combating climate change impacts involves the use of research, indigenous knowledge, effective policies, and people's participation in strengthening the communities and minimizing the adverse effects.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

Similar to all other studies, this one has a unique methodology. To investigate the effect of average temperature and rainfall on forest areas throughout time, a systematic and thorough technique was used. Also which model fits well to predict temperature data was one of the main goals of this study. The study began with a clear goal: to better understand the relationship between climatic variables and forest cover patterns. Comprehensive data were obtained from credible sources, including historical records of average temperature, rainfall, and forest area over time. By creating a UI interface the model was implemented to predict temperature on a year and month basis.

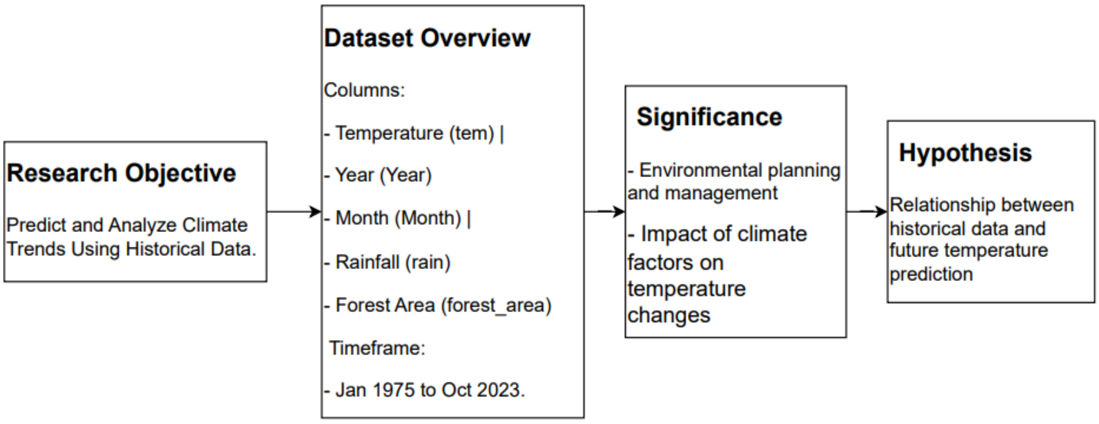


Figure 3.1.1: Working Procedure of Methodology

The following Figure 3.1.1 illustrates the working procedure of the methodology, along with a synopsis of the entire research process.

The purpose of this study is to forecast future temperature levels based on past climatic variables and the area of forested lands. The date used consists of data ranging from January 1975 to October 2023 with the inclusion of variables that are designed as Temperature, Year, Month, rainfall, and Forest area. Such a rich database enables one to build a stable ground for climatic long-term change analysis and temperature fluctuations therein. The importance of this study is based on the prospect that it would contribute to improving environmental planning and management since it will provide estimations about further tendencies that would be useful when making decisions regarding the allocation of resources. Since the climate and rainfall forecasts involve climatic comparisons with the history of weather and forest areas, the study will be planning a relationship between these factors and expected future temperatures. Such forecasts are useful in the prevention of the impacts of climate change variability and in the promotion of sustainable and proper use of the environment. The hypothesis on which this research is based postulates that analysis of historical data containing temperature, rainfall, and area of forest can help in predicting change in temperature in the future, which is an important instrument in the framework of climate change management and adaptation processes.

3.2 Research Subjects and Instrumentation

"Using Machine Learning to Predict and Analyze Climatic Trends and: A Data-Driven Approach" is the title of the thesis. The purpose of this study was to determine how rising temperatures, rainfall, and deforestation are related. Several trustworthy internet sites were used to gather the data. To make this research possible, it was necessary to read through a few other publications that discuss climate change, deforestation, global warming, forest area analysis, and weather forecasting data analysis. Setting the goal and gathering the data for this study was the most difficult aspect. Since the source must also be verified, it was exceedingly difficult to find the weather and forest area data of underdeveloped nations like Bangladesh. This study emphasized the importance of both hardware and software. The tools and technologies that helped to come up with the conclusion are given below:

Software and hardware:

Table 1: Hardware Software Requirements and Development Tools

Hardware and Software Requirements	**Development Tools**
Intel Core i5 8th generation with 8 GB of RAM	Visual Studio Code
1TB HDD	Python (Pandas, Seaborn, NumPy, Matplotlib, Sklearrn)
Google Spreadsheet	Flask
Microsoft Excel 2016	HTML

Table 1 shows the Hardware Software requirements of the study and required development tools for the study.

3.3 Data Collection Procedure

The toughest part of this research was collecting data. The dataset contained several data types: average rainfall, average temperature, and average forest area. Syncing the timeline and location was the biggest challenge while conducting this study. In this study, data was collected from numerous online sources. NASA's POWER Project provides the temperature data, while the Humanitarian Data Exchange (HDX) website provides the rainfall data, with particular attention to Bangladeshi rainfall data [8]. The forest area data was collected from World Bank Data on Forest Area and FAO Bangladesh Country Paper [9][10]. There are four types of data present in this dataset. These are Temperature(C), Date, Rain (millimeters), and Forest Area(percentage). 586 sets of data are available in this dataset. In this research, the main propaganda was to find out the relationship between deforestation and the drastic increase in temperature and the major change in the pattern of weather. With this dataset, it is possible to study the temperature, rainfall, and forest cover from January 1975 to October 2023.

Table 2: Sample Dataset

Average Temperature	**Month**	**Year**	**Average Rainfall**	**Forest Area**
18.0454	January	1975	2.66038	15.48
20.7283	February	1975	12.4828	15.48
25.4553	March	1975	12.8183	15.48
28.2002	April	1975	94.5976	15.48
28.0144	May	1975	251.874	15.48
28.3837	June	1975	338.62	15.48
27.1209	July	1975	682.34	15.48
27.9833	August	1975	279.506	15.48

The given table 2 is the visualization of sample datasets containing average temperature, month, year, average rainfall, and forest area, respectively. While the average rainfall data is derived from the Humanitarian Data Exchange (HDX) website, with a particular focus on Bangladesh rainfall data, the average temperature data is taken from NASA's POWER Project [9]. Get information on the amount of forest cover by obtaining World Bank Forest Area Data and FAO Bangladesh Country Paper [10][11].

3.3 Data Preprocessing

Data preprocessing is required to get the best result from bias-free, well-formed data. To perform this research, several data sources were used to collect well-formed data that will properly sync the timeline and location of the variables like average temperature, average rainfall, and average forest area respectively. To get the best version of the collected data few preprocessing steps were performed, inspect the dataset, handle missing data, encode categorical variables, standardize numerical features, check for outliers, and split the dataset. Here is Figure 3.3.1 illustrates the steps of how the data preprocessed.

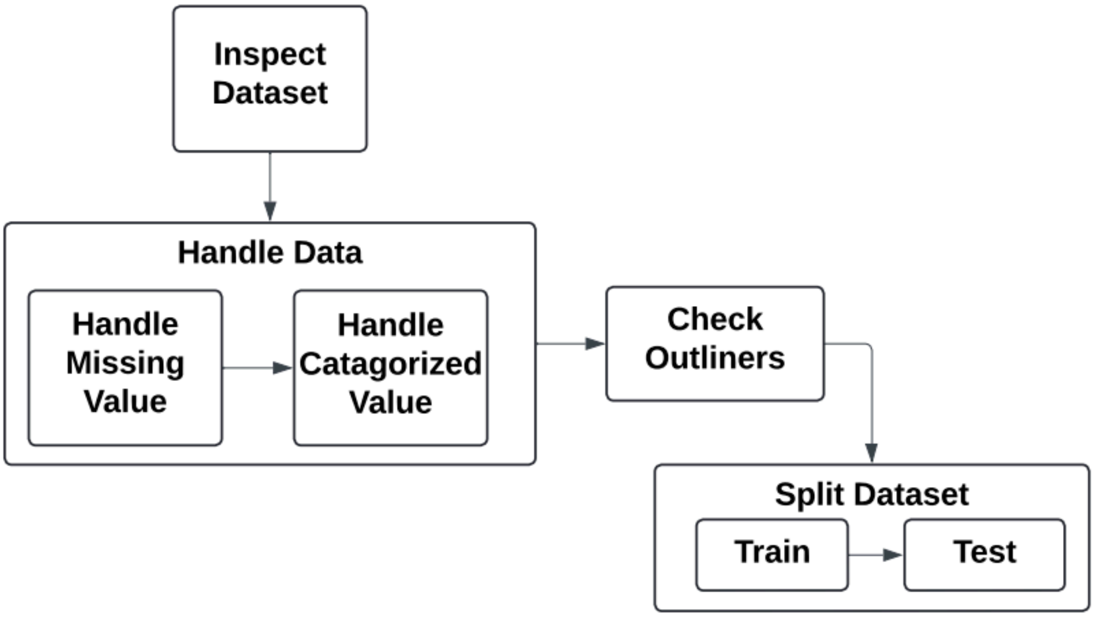


Figure 3.3.1: Data Preprocessing Diagram

Inspect Dataset: Understanding the dataset's quality and composition is essential for study because it offers information. One way to prevent biased outcomes is to identify missing numbers early on and employ suitable handling procedures. Analytical techniques and precise conversions are guaranteed by an understanding of data types, such as numerical, categorical, and date time. An examination of the dataset's overall structure can reveal information about its organization, complexity, and existence of nested or hierarchical linkages. Making plans for the next steps, spotting anomalies, and guaranteeing data integrity are all aided by this inspection. A thorough inspection of the dataset is necessary for precise analysis, to minimize possible issues, and to boost the dependability of the findings. [32]

Handle Missing Value: There may be missing data at the attribute (item) level or the case (person) level. The complete case analysis (CCA) or "exclude by listwise" approach is the standard for handling missing values. In survey-based research, missing data must be considered throughout the entire process, from design to analysis. [10] Data was gathered from multiple data sources, so it is necessary to determine whether any given value is null or not. The next step is to handle them with the proper approach.

Handle Categorized Variables: To create data, managing categorical values is essential. Utilizing encoding techniques, categorical data must be represented numerically. If an ordinal connection is missing, binary columns can be created for each category using one-hot encoding. The dataset's dimensionality may rise as a result, though. Despite its space-saving benefits, label encoding places integers in categories, which could unintentionally result in ordinal linkages. Good management ensures accurate analysis and unaltered data. [33]

Check Outliers: To properly prepare data, it is imperative to check for outliers, as these can greatly distort analysis and model performance. Several things can lead to outliers, including measurement errors, incorrect data entry, and true but uncommon events. To find these abnormalities, statistical techniques like the Z-score and Interquartile Range (IQR) are frequently applied [32]. While the Z-score approach identifies values that are many standard deviations away from the mean, the IQR method highlights data points that lie beyond 1.5 times the interquartile range. Through the identification and management of anomalies, such as elimination or modification, investigators can guarantee a more precise and resilient examination.

Split Dataset: Separating the clean dataset into training and testing sets is a crucial step in confirming the model's effectiveness. While the training set is used to train the model, the testing set evaluates the model's accuracy and suitability for new data. This process helps avoid overfitting, a condition when a model performs well on training data but poorly on unidentified data. A common split ratio is 70% for training and 30% for testing, though this could vary based on the size of the dataset and specific research needs [33]. Properly splitting the dataset ensures the model's robustness and reliability and provides a clear picture of its prediction performance.

3.4 Statistical Analysis

Data were gathered for this study from a variety of sources. Here, historical data is used to comprehend how deforestation and climate change are related. Here, average temperature, average rainfall, and average forest area were utilized to calculate the output. The specified variables' historical data was applied between 1975 and 2023. The collection includes 48 years of historical data to assess the effects of extreme heat waves, the rate of deforestation, the relationship between temperature and rainfall, and many other topics. From a statistical standpoint, various conclusions can be made from this dataset, which contains temperature, rainfall, month, year, and forest area.

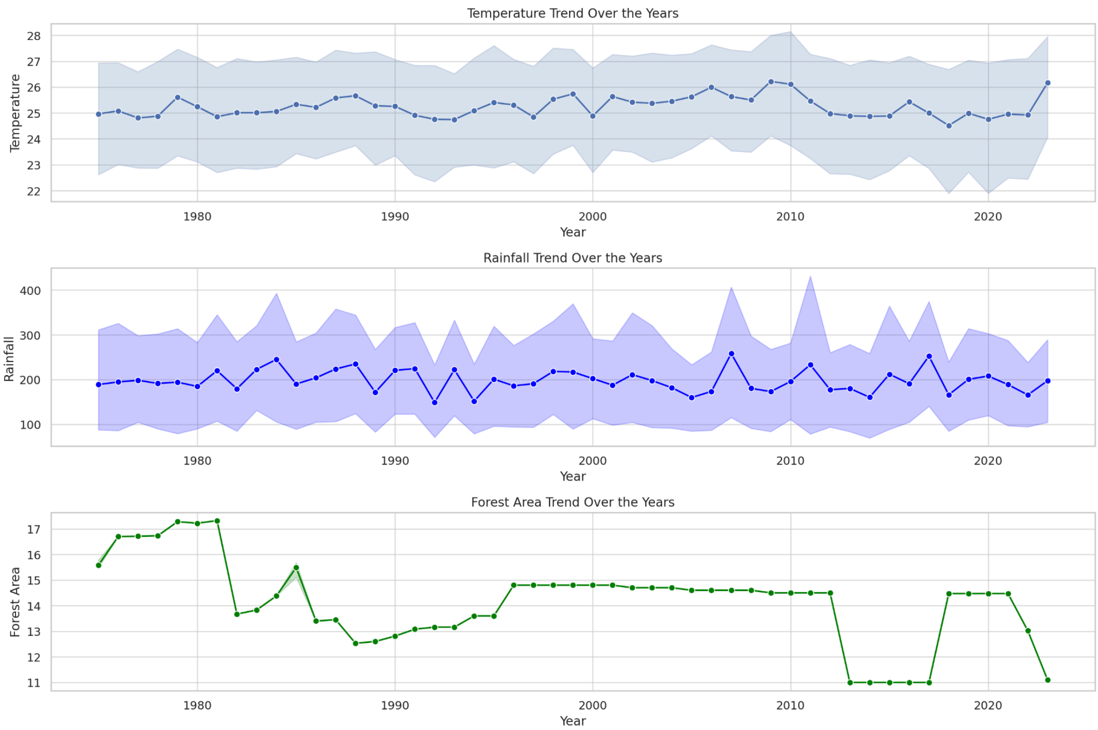


Figure 3.4.1: Temperature, Rainfall, Forest Area Trend Over the Years

In Figure 3.4.1 we can visualize the temperature, rainfall, and forest area trend over the year.

Temperature data shows a stable positive trend which implies a warming climate trend. Monthly temperature breakdown highlights a clear seasonal cycle wherein temperatures are higher in summer than at any other time whereas they are lower in the winter months. There are always seasonal fluctuations in climate that are well within the standard behavior of temperate zones. However, for the last couple of years, average temperatures seem quite high compared to previous historical data. Maximum temperature shows between 30 degrees to 40 degrees Celsius in summer making April the warmest month of the year. [34]

Precipitation data exhibits a high degree of variance at both inter-annual and intra-annual levels. The variability seen here may therefore reflect the fact that we've yet to isolate what caused the spikes—some things are seasonal, some change every year, and others might have broader climatic influences at work, such as El Niño or La Niña. The great positive relationship between temperature and precipitation indicates that greater temperatures generally coexist with more rain, therefore representing a monsoonal or warm weather pattern.

The time series of forest area shows one of high stability in this region implying that the effective forest management characteristics in this region were until now fully developed or this forest ecosystem suffered low changes during this period. The weak to non-existent correlations between forest area and temperature and rainfall would suggest that the forest area has not been strongly influenced by inter-annual changes in the local climate, while longer-term effects or regional signals are still a possibility as are other environmental drivers.

Overall, the Compro data set is highly valuable as it can be utilized to determine climatic patterns and analyze hypotheses for impacted forest plan areas. It also highlights the key patterns and connections needed to understand how ecological and environmental dynamics will evolve.

3.5 Proposed Methodology

Using historical datasets as a basis, this study aims to comprehend the regional consequences of climatic trends and the relationship between temperature, rainfall, and forest area and their roles in warning of global heat waves. This analysis was carried out using 48 years' worth of climate and forest area data. They are understanding the data through exploratory data analysis allowed for the automation of some tasks using comparative machine learning models. The proposed methodology is depicted in Figure 3.5.1.

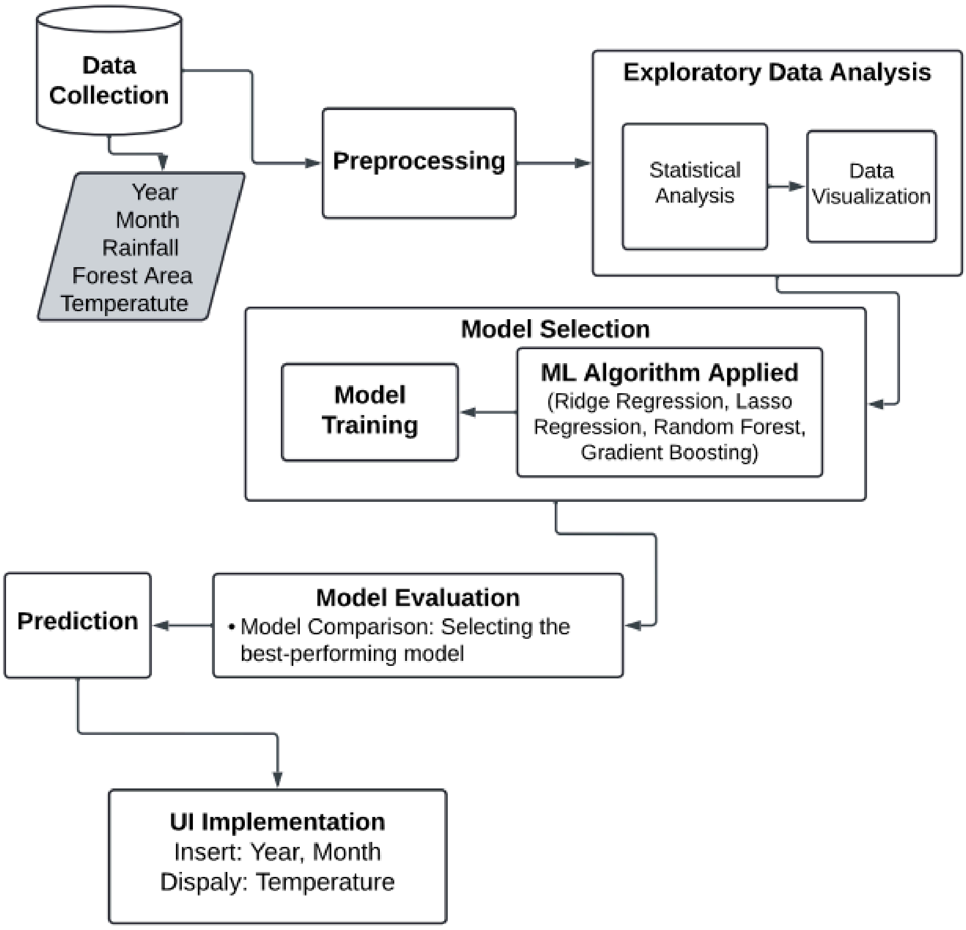


Figure 3.5.1: Proposed Methodology Diagram

3.5.1 Data Collection

Gathering data was the most difficult aspect of this research. Numerous distinct forms of data are used in the dataset. Here, the most difficult aspect was coordinating the dataset's timeline with the location. A variety of internet sources provided the data. The used dataset includes historical information on rainfall, temperature, and forest area and consists of 586 pieces of data in all. To complete this analysis, 48 years' worth of data—from 1975 to 2023—were collected.

3.5.2 Data Preprocessing

To obtain a better set of data that is well-formed and free of bias, data preprocessing is an essential component of every research project. Rearranging the data was necessary because the collected data came from several sources. Subsequently, examining the dataset, managing absent values, managing classified variables, verifying anomalies, and segmenting the dataset facilitate the acquisition of a well-organized version of the gathered dataset.

3.5.3 Exploratory Data Analysis (EDA)

Exploratory data analysis, or EDA, is a well-established statistical discipline that provides theoretical and numerical methods for pattern recognition to aid in the creation and refinement of hypotheses.[35] To better comprehend the data and display the links within the data, statistical analysis, and data visualization were carried out utilizing exploratory data analysis approaches. In the visualization part Plot time series of temperature, rainfall, and forest area to display trends and scatter plots are being used to assess correlations between variables. In the statistical analysis portion estimate the mean, median, variance, and other statistics for each variable and explore relationships between temperature, rainfall, and forest area.

3.5.4 Model Selection

Here we conducted a comparison analysis using four distinct machine learning models. Models that were chosen for the analysis include Random Forest, Gradient Boosting, Ridge Regression, and Lasso Regression. 99.46%, 99.42%, 20.96%, and 18.29%, respectively, are the accuracy percentages of the chosen models. In light of this, the random forest model outperforms the other four models when comparing the results.

Random Forest: Random Forest is an ensemble learning method that constructs a multitude of decision trees during training and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees. It uses bagging (Bootstrap Aggregating) to create subsets of data, train decision trees on these subsets, and then aggregate their predictions.

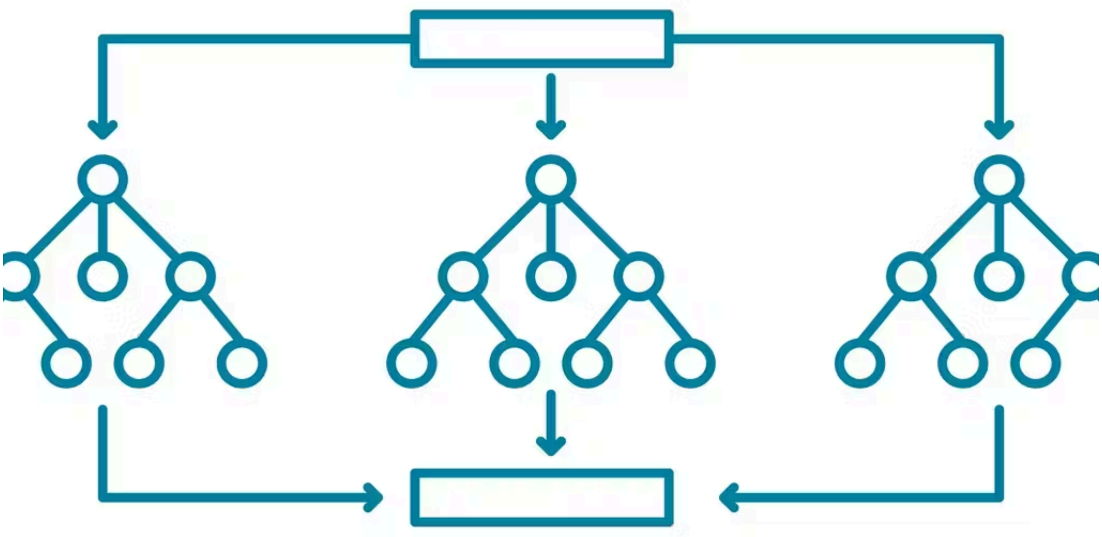


Figure 3.5.4.1: Random Forest [41]

By Figure 3.5.4.1 we can visualize the structure of the Random Forest Model.

Gradient Boosting: Gradient Boosting is another ensemble learning technique that builds models sequentially. Each model attempts to correct the errors of the previous model. It combines weak learners (usually decision trees) into a strong learner in an iterative fashion. The key idea is to minimize the loss function by adding new models that predict the residuals or errors of prior models.

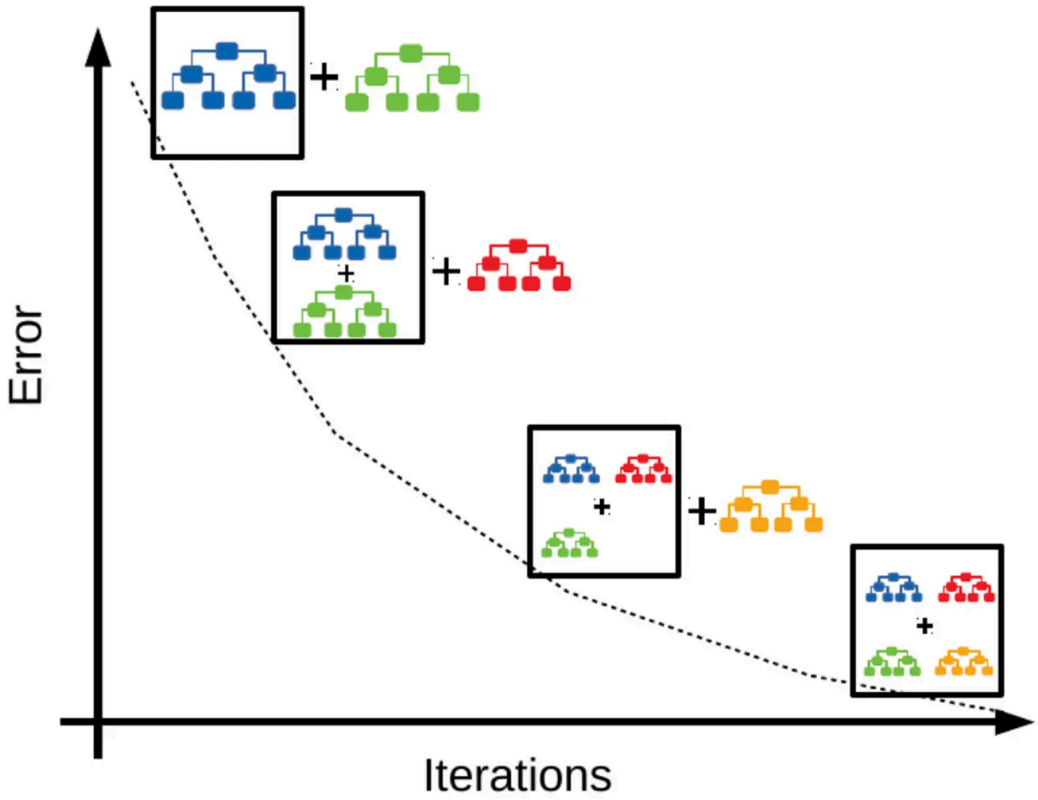


Figure 3.5.4.2: Gradient Boosting [42]

By Figure 3.5.42 we can visualize the structure of the Gradient Boosting Model.

Ridge Regression: Regression is a type of linear regression that includes a regularization term (L2 regularization). This regularization term penalizes large coefficients to prevent overfitting. The objective function is modified to include the sum of the squared values of the coefficients, which helps to keep them small and reduces model complexity.

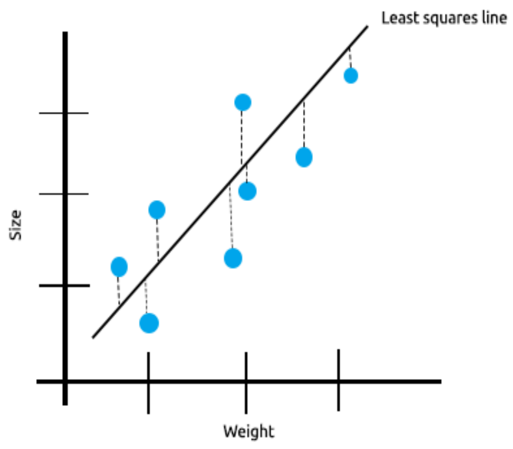


Figure 3.5.4.3: Ridge Regression [44]

By Figure 3.5.4.3 we can visualize the structure of the Ridge regression Model.

Lasso Regression: Lasso Regression (Least Absolute Shrinkage and Selection Operator) is similar to Ridge Regression but uses L1 regularization. This regularization term penalizes the absolute values of the coefficients, which can result in some coefficients being exactly zero. This makes Lasso Regression useful for feature selection as it can shrink coefficients of less important features to zero.

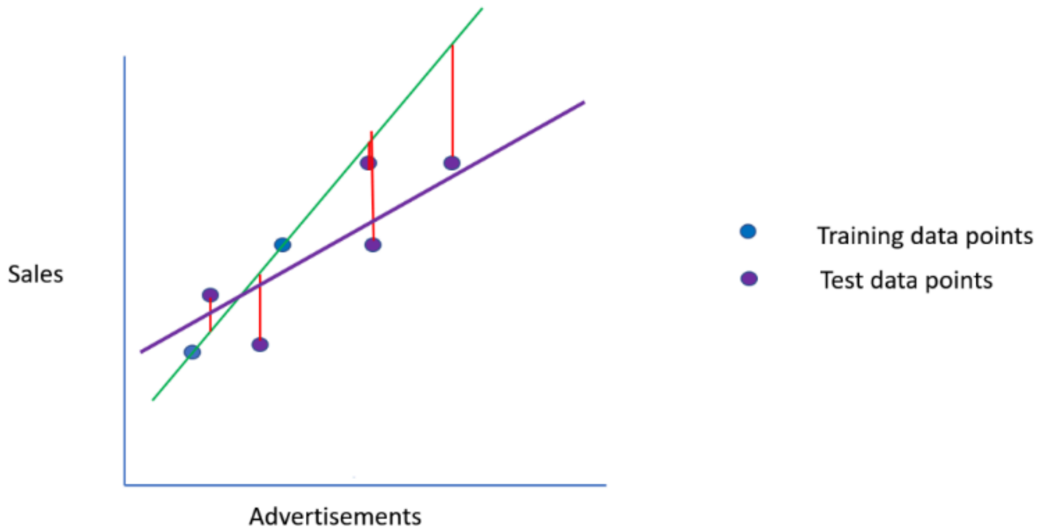


Figure 3.5.4.4: Lasso Regression [43]

By Figure 3.5.4.1 we can visualize the structure of the lasso Regression Model.

3.5.5 Model Evaluation

R-squared metrics were used in a thorough analysis of the model's accuracy. These indicators were selected to provide a clear picture of how effectively each model anticipated changes in the amount of forest based on fluctuations in rainfall and temperature. R-squared provided a gauge of the model's ability to account for the variation in the forest area data. The performance of each model might be meaningfully quantified by using these metrics. Finding the model that produced the most accurate rainfall coincides with the end of heat waves. Data on climate does not appear to be impacted by forest areas. It appears that the four reliable methods that have been applied to generate several machine-learning models have also been successful. Robust methods are required to complete this step. For forecasting changes in forest areas based on climatic variables, the model that maximized the R-squared value was thought to be the most accurate. This extensive assessment guaranteed the final model's robustness and dependability, which were crucial for insights and decision-making.

3.5.6 Prediction

The experimental analysis includes four models, namely, Ridge Regression, Lasso Regression, Random Forest, and Gradient Boosting to compare which outperforms the other model on the temperature data to be used in forecasting future temperatures. This pattern is learned from the available data of the temperature, year, month, rainfall, and area of the forest from January 1975 to October 2023. The validity and efficiency of the models is compared through erogenous coefficients of determination (R^2). Thus, the model that has the highest efficacy levels received after the mentioned evaluations is used for the prediction. The selected model, therefore, was applied in identifying future temperatures for the period 2024 to 2036. Since the trends have been established for historical data, it is possible to make estimates for the temperature change not only within the range of the created dataset but also after a certain number of years in the future. In making future forecasts, the model requires values for rainfall and forest area for future years. These values can be worked out either based on historical data or from other models or assumptions if made.

3.5.7 UI Implementation

With employing the Flask, a lightweight web frame for Python as the UI, the user interface (UI) is applied. Python implementation of Flask has made it easier to incorporate the back-end code written in Python with the front-end components of web-based systems. For actual use, a basic HTML form is developed to accept year and month from the user to predict the temperature. It consists of input boxes for the year and the month as well as the execute button for the prediction of the rate. Home Route: Shows the HTML form to the user. Predict Route: Responsible for capturing the form submission, reading the values inputted by the user, and working on the backend model. In the next year and month that is inputted by the user, the backend retrieves the other necessary features of rainfall and forest area. The model that is then used to estimate the temperature anticipated in the given year and month is then obtained. Thus, the predicted temperature is returned to the user through the web interface and they can see the immediate results of the model's calculations. The UI ensures that the prediction process is user-friendly and accessible, making it easy for users to interact with the prediction model.

3.5.8 Interpret result

It is evident from evaluating the statistical analysis's outcome that there are significant relationships between temperature and rain. It appears most likely that periods of high odds with the accuracy of 96.01%, 95.872%, 94.07%, and 93.89%, respectively, are Gradient Boosting, Random Forest, Ridge Regression, and Lasso Regression,. Out of all the models, the gradient-boosting model produces the best results.

3.6 Implementation Requirements

An application of a model for practical work necessitates its implementation. To implement the model you must fulfill several crucial conditions before you can put a model for assessing climate and forest area data into practice. Gather and prepare the data first, making sure to handle any missing values and format them correctly. For data analysis and model building, use Python and tools such as pandas, numpy, matplotlib, seaborn, and sci-kit-learn. Sufficient hardware is required, such as a PC with a minimum of 8GB of RAM and enough storage. Divide the data into training and testing sets after preprocessing it by engineering pertinent features and normalizing it. Create and train models, such as time series forecasting's ARIMA or trends' linear regression, then assess their effectiveness with the right metrics. Analyze the data visually to identify patterns and connections. Lastly, test the model in a real-world setting and keep an eye on its performance. Make sure to record every step for accuracy and repeatability.

CHAPTER 4
EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

This is clearly illustrated in this text where the author first presents four machine-learning models before going about the chapter to identify the best model. The Random Forest, Gradient Boosting, Ridge Regression, and Lasso Regression are chosen models of the model's type. Random Forest forms many decision trees to help it identify the relationship between variables more easily. Gradient boosting aims at creating trees one by one after the other with more focus on correcting the mistakes made by the previous tree. Through multicollinearity and regularization, the Ridge Regression can help overcome the possible problem of model overfitting. Apart from modulating the frequently observed problem of overfitting through the process of regularization, Lasso Regression also selects those features that it agrees with by setting a part of its parameters to zero. The raw data was compiled from various online sources and the data in this paper encompasses historical information on temperature, rainfall as well as forest area. The exploratory data analysis section helps in understanding the data narrative and the visualization part. For running the model VS code is used. After model evaluation by using the model with the best accuracy, a UI interface was implemented to predict future climate trends.

4.2 Experimental Results

The data collected must be analyzed in various experiments to confirm or deny the theories formulated and new knowledge gained. To make sure the results are meaningful and reliable, the first steps involve data collection and data cleaning. The final step is statistical analysis commonly known as inferential and descriptive analysis to present and analyze the data collected. Some important patterns are distinguished and it becomes easier to analyze complicated data using tools such as graphs and charts. These are basic steps of interpretation where the obtained outcomes are compared to the hypothesis towards categorizing trends, admitting research limitations in the process. The conclusion involves summarizing the main findings, discussing the implications of the study on a broader scale, and suggesting potential areas of study for future research. Finally, verification of the results and dissemination to the scientific fraternity involve the preparation of written documents such as reports or scientific papers in formats that can be peer-reviewed and the use of visual aids. By employing exploratory data analysis and model evaluation, were able to conclude the experimental result in this study.

Exploratory Data Analysis (EDA)

Exploratory data analysis which is abbreviated as EDA is the procedure of inferring about the nature of the data by the help of graphical tools and the help of descriptive statistics [38]. Exploratory data analysis is defined as the sort of knowledge that looks at the data from its best side. Exploratory data analysis is to get the maximum information about the variable in question. Different statistical techniques and data graphical techniques assist one in getting a better understanding of the data. In order to comprehend the data story numerous analysis techniques such as data visualization and descriptive data analysis have been used in this field of study.

Statistical analysis

Without a doubt, one of the most sophisticated, versatile, and often utilized tools in the field of sensory analysis is descriptive analysis [36]. Descriptive analytics is a subset of data analytics that describes the current situation using information from the past [37]. In this historical data-driven investigation, the desired outcome was obtained by utilizing 48 years of temperature, precipitation, and forest area data. You may see the dataset's statistical analysis, including the quantity of data, mean, standard deviation, minimum, maximum, 25th, 50th, 75th, and 100th percentiles, by referring to Table 2. Below is an explanation of the table.

Table 3:: Descriptive Analysis Summary

Statistic	Temperature (tem)	Month	Year	Rain	Forest Area
Count	586	586	586	586	586
Mean	25.24	6.48	1998.92	197.45	14.15
Std	3.69	3.45	14.11	189.6	1.63
Min	15.93	1	1975	0.00089	11
25%	22.45	3.25	1987	20.61	13.16
50%	27.13	6	1999	162.58	14.5
75%	28.16	9	2011	345.05	14.8
Max	29.53	12	2023	1012.02	17.32

Temperature, month, year, rain, and forest area are the five categories of data that make up the dataset, as shown in Table 3. There is not a single null value in any of the 586 data points in the dataset. You can see that there is moderate variability by looking at the column temperature, where the mean temperature is 25.24 and the standard deviation is 3.69. A mean of 1998.92 and a standard deviation of 14.11 are found for the years 1975 to 2023. The amount of rainfall varies greatly; it ranges from almost zero to 1012.02 units, with an average of 197.45 units and a standard deviation of 189.60 units. The rainfall ranges from 0.00089, nearly nil, at the minimum to 1012.02, quite a big quantity, at the maximum. The range of the forest area is 11 to 17.32 units, with a mean of 14.15 and a standard deviation of 1.63. This suggests that forests hold their characteristics for a very long time. The maximum forest area, which is 6.32%, makes it evident that there is a significant gap between the lowest and maximum forest areas. Such a high rate of deforestation must be concerning in a small nation like Bangladesh.

Data Visualization

Data visualization is necessary for the scientific method to be complete. With the use of powerful visualizations, a scientist may understand their data and communicate their discoveries to others [39]. When most visualization techniques were developed, it was assumed that the data they would be used to portray were error-free. But things aren't always like this. In the visualization field, adding a signal of uncertainty to visual representations has become a new endeavor [40]. The correlation of the data visualization approach has been used in this study to comprehend and illustrate the yearly patterns of data. The relationship between each dataset variable is shown in Figure 4.2.1. The month-to-month trend by temperature and rainfall is shown in Figures 4.2.3 and 4.2.2.

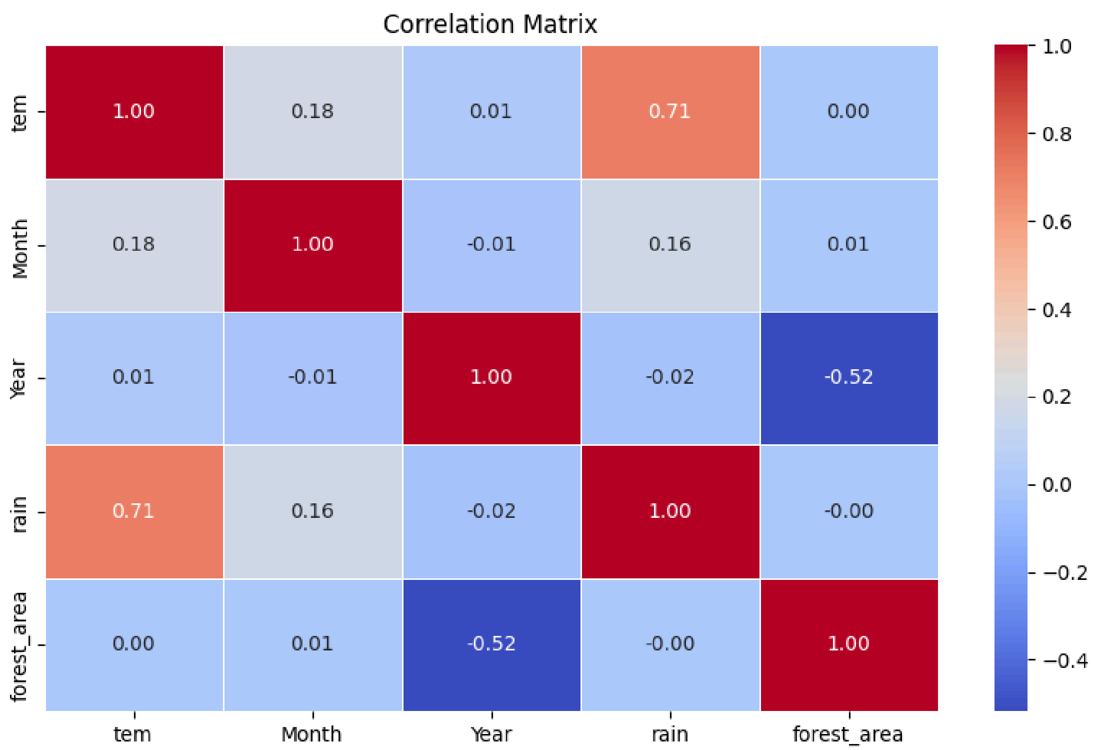


Figure 4.2.1 Correlation Matrix

By observing Figure 4.2.1 it is clear that there are highly positive correlation between rainfall and temperature. That means if the temperature increases, the rate of rainfall is also looking upward. Here the correlation between rainfall and month is 0.71. This could be observed in certain tropical or monsoon climates where higher temperatures lead to higher rates of evaporation and subsequent precipitation. On the other hand, the correlation between year and forest area is -0.52. This suggests that there is a moderately negative connection between the forest area and the year, meaning that the quantity of forest area has been declining over time at a steady but not particularly fast rate. According to the correlation, a decrease in forest area may have a major negative impact on the environment, such as a reduction in carbon sequestration capacity, a disturbance of ecosystems, a loss of biodiversity, and an increase in greenhouse gas emissions. The correlation between other variables are almost zero, which means there is not that much connection between other variables.

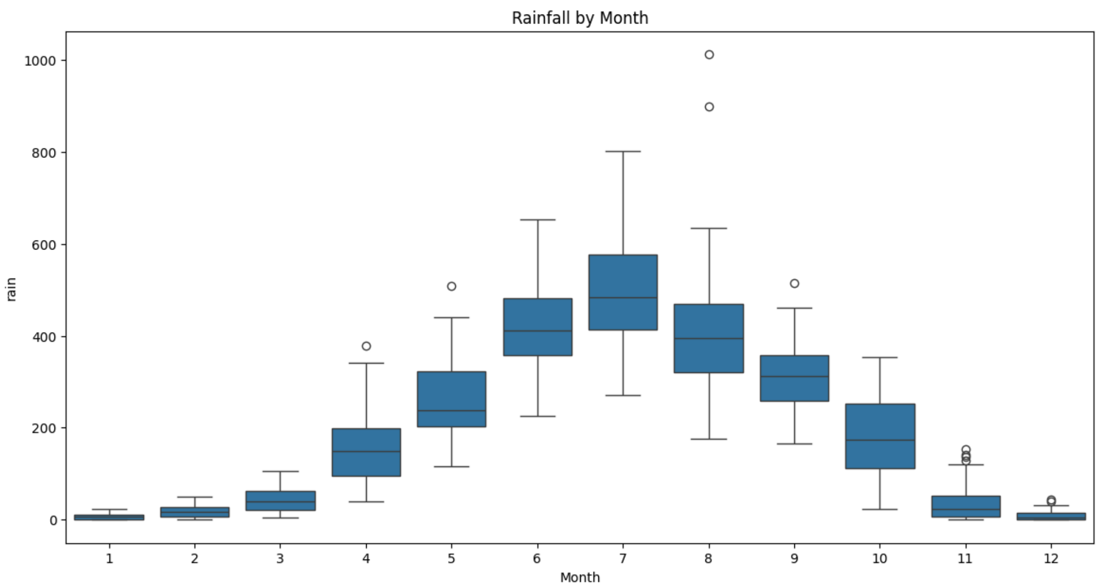


Figure 4.2.2: Rainfall by Month

A closer look at Figure 4.2.2 leads us to the conclusion that July is the wettest month of the year. Rainfall ranges from 250 to 800 millimeters each month on average. At the median, it is nearly 450. Winter has the least quantity of precipitation because there is almost no rainfall throughout this season. The hyperbolic structure is shown in the rainfall-by-month graph, which means that the annual rainfall increases and lowers throughout the year, peaking in the middle and falling at the beginning and end.

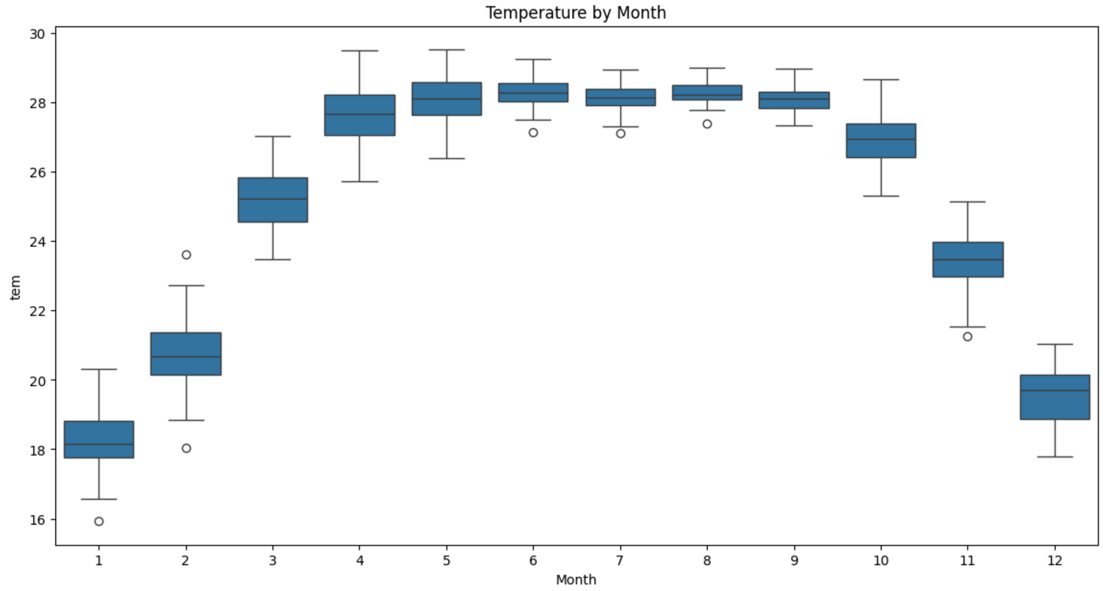


Figure 4.2.3: Temperature by Month

Figure 4.2.3 depicts the monthly temperature trends, revealing a distinct summer bias throughout the year. Six months are characterized by notably high temperatures, while winter holds sway for only two months—December and January. These winter months provide a brief respite from the pervasive heat, marking the only period when cooler weather prevails. Spring offers a moderate climate, serving as a pleasant transition between the temperature extremes. From April to October, temperatures consistently range between 26 and 30 degrees Celsius, with May standing out as the hottest month of the year, and January as the coldest. This prolonged heat wave indicates a significant climatic issue. Even the rainy season, autumn, and late autumn are dominated by hot weather, suggesting a disruption in the natural weather patterns. This persistent heat, extending across typically cooler seasons, underscores the region's struggle with elevated temperatures year-round.

Model Evaluation

Four machine learning models were used here for random forest, gradient boosting, ridge regression, and lasso regression. Then predictions and evaluations were made by their performance using R^2.

Table 4: Accuracy of Different Models

Models	Accuracy
Gradient Boosting	96.01%
Random Forest	95.57%
Ridge Regression	94.07%
Lasso Regression	93.89%

From Table 4, we can visualize the accuracy of different models that were developed by using various regression models based on the R^2 values converted to percentages.

Gradient Boosting: Additionally, the Gradient Boosting model has remarkable performance, with an accuracy of 96.01%%. It is a very dependable predictor because, like the Random Forest model, it accounts for nearly all of the variance in the forest area.

Random Forest: With an accuracy of 95.57%, the Random Forest model explains a great deal of the volatility in the forest area. This suggests that the model is highly successful in accurately predicting outcomes and identifying underlying patterns in the data.

Ridge Regression: The forest area's variance is explained by the Ridge Regression model to the extent of 94.07%. This suggests a comparatively inadequate match in contrast to the Random Forest and Gradient Boosting models. For this dataset, Ridge Regression is not very effective, however, it is marginally better than Lasso Regression.

Lasso Regression: Merely 93.89% of the variance in the forest area can be explained using the Lasso Regression model. Given that the models evaluated had the lowest accuracy, Lasso Regression is not a good fit for this dataset.

Insights

Non-linear Models: With accuracy rates over 95%, both Random Forest and Gradient Boosting, ensemble techniques that can represent intricate non-linear relationships, exhibit exceptional performance. This implies that the complicated and non-linear interactions in the data are probably what these models are best at capturing.

Linear Models: The Ridge and Lasso Regression are examples of linear models that have superior accuracy levels above 93%. This suggests that linear models also work quite successfully with this dataset.

Model Selection: Although both linear and non-linear models perform great in this case. However, non-linear models like Random Forest and Gradient Boosting are slightly better than linear models like ridge regression and lasso regression. So for moving forward non-linear models will be the best options for this dataset.

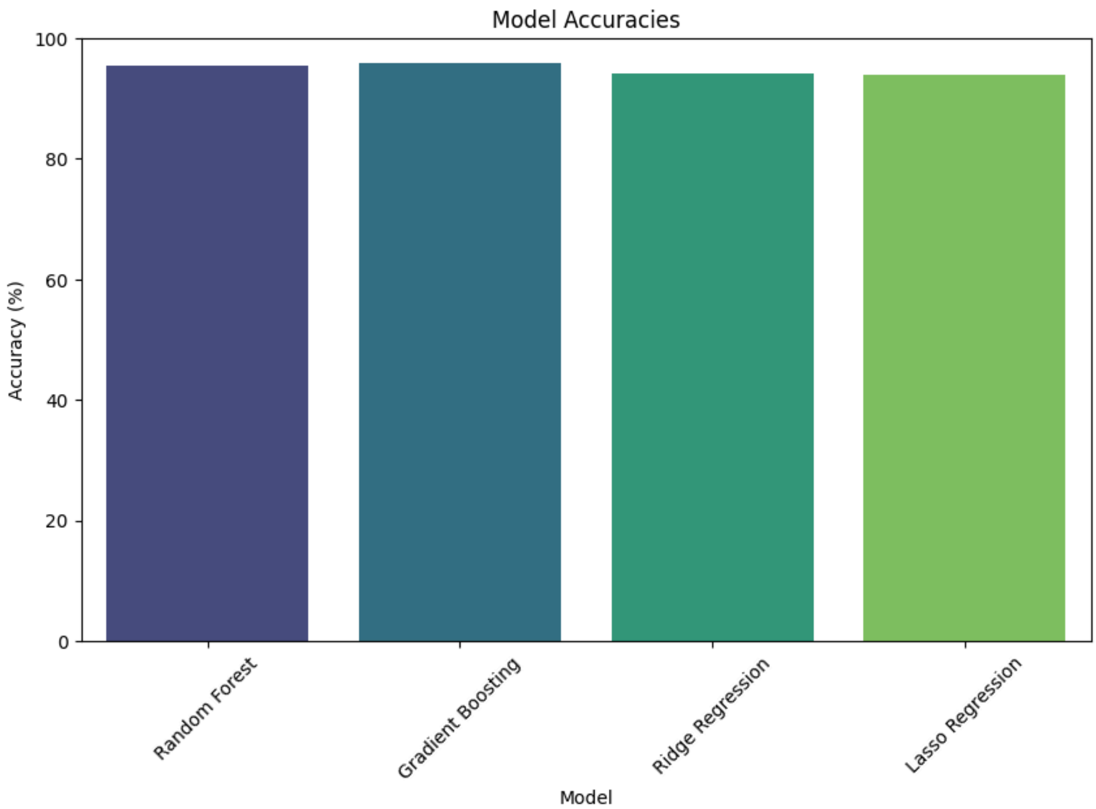


Figure 4.2.4: Model Accuracy Comparison

This visualization through Figure 4.2.4 will clearly show the difference in accuracies between the various models, highlighting the effectiveness of Random Forest and Gradient Boosting compared to Ridge and Lasso Regression.

Predict Future Temperature

The future temperature is predicted by using the historical data ranging from January 1975 to October 2023 including temperature, year, month, rainfall, and forest area. Using more complex models like Ridge Regression, Lasso Regression, Random Forest, and Gradient Boosting on this dataset, we derived the model that yielded high accuracy in predicting the temperatures. To enhance the system's perceived usability and provide a means for users to get predicted temperatures, we employed a Flask web application wherein users can enter a specific year and month and get the latter as the response. This system downloads the rest of the additional features, namely rainfall and size of the forest, from history to be able to provide competitive forecasts for future dates. This approach not only shows the ability to predict climate with the help of machine learning but also gives practical possibilities to apply them and manage the environment.

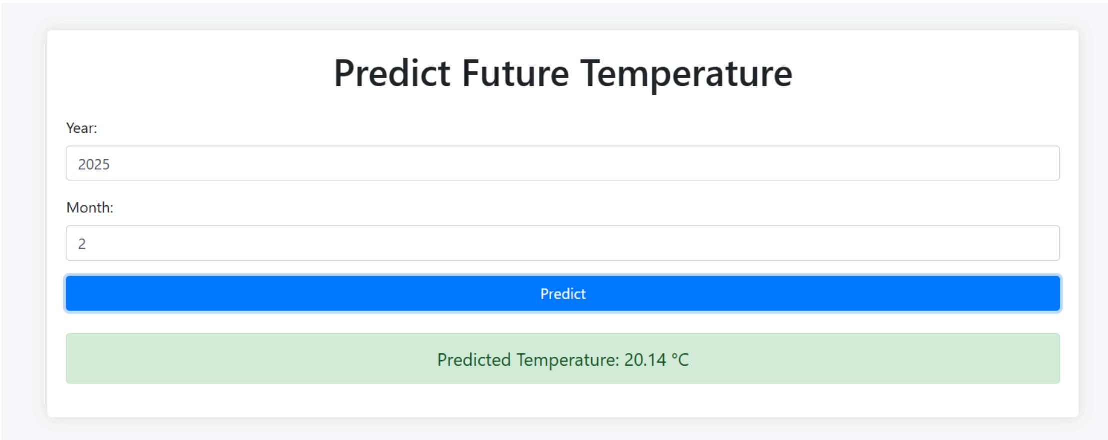


Figure 4.2.5: UI of Predict Future Temperature

From Figure 4.2.5, we can visualize the user interface of the application which was made based on the dataset and the best-fitted model and the interface also shows input and output of one predicted monthly average temperature.

Table 5: Predicted Future Monthly Average Temperature

Year	**Month**	**Predicted Temperature**
2025	February	20.14°C
2032	September	24.87°C
2029	December	19.49°C
2024	October	24.51°C
2036	January	17.97°C

The provided Table 5, which was created using the best-fitted model Gradient Boosting to produce the prediction model's user interface, shows the expected monthly average temperature data using the developed user interface from the climate trend dataset. By looking at the table it is possible to predict that the average temperature in 2025, February could be 20.14°C. In parallel, the table shows 2032, September's monthly average predictable temperature will be 24.87°C. For December 2029, October 2024, and January 2036 predictable temperature is 19.49°C, 24.51°C, and 17.97°C respectively.

4.3 Discussion

Analyzing the results of the several regression models that were used to forecast forest area, it was found that there was significant variation in the efficacy of the different models. The equation-fitting models obtained presented very high R2 values of 96. 01% and 95. 57% and 57% correspondingly, non-linear ensemble models—Gradient Forest and Random Boosting presented extremely high percent of accuracy. Altogether, because they can accurately capture the structural complexity of many forms of non-linear relationships in the data, these models are good predictors. However, with accuracies of 94. 07% and 93. 88%, and 89% respectively while the other linear models such as the Ridge and Lasso Regression are also good. This comparison shows that both non-linear modeling and linear modeling are sufficient to depict the intricate relation in the depiction of the given data set. The relationships in the data may be complex and there may be different type of relations, for which many linear models as well as non-linear models can be very helpful. However, the non-linear models behave slightly better than the linear models According to them Gradient boosting as a non-linear model has the highest performance score by owning 96. Hence, Gradient Boosting is the preferable model for the occasion and, therefore, it is advised to work further on the given dataset with Gradient Boosting as a model of 01% accuracy. It is for the implementation of the predictive analysis that gradient boosting is used. The above-stated studies support the need to choose appropriate modeling methods that reflect the basic architecture of interactions concerning the simplicity and non-linearity of the data patterns.

CHAPTER 5
IMPACT ON SOCIETY, ENVIRONMENT, AND SUSTAINABILITY

5.1 Impact on Society

Thus, the given work has many social implications. The conclusions contribute to enhancing people’s awareness of the utilities of the forests and how climate change impacts the vital systems. This might increase public support for the conservation strategies and ecological appropriate behaviours in future. The research findings in this study can further help curricula at educational institutions in a bid to teach the next generation on the importance of forests and the need to address climate change. These results may be of great use to the forestry sector as it will enable the making of long-term procedures that will reduce the threat of leading to resource depletion. Forests management can increase the ecotourism since more tourists will be attracted by the preserved environment and their love for it.

5.2 Impact on the Environment

This study’s relevance to the environment is also high. The knowledge of climatic factors influencing forest areas can therefore help in the enhancement of the techniques in forest management. Thus, reforestation, afforestation, and techniques of forest conservation might be less arduous and more efficient in due course, thereby preserving biodiversity and ecological services. It must be mentioned that today, a significant portion of the populations of all species on Earth, inhabiting the land, can be met in the forests. The research findings can be useful in maintaining the ecosystem by identifying areas that need to be protected from the unfavorable impacts of climate change. Thus, protecting forests means protecting our planetary air-conditioning system: sequestering carbon dioxide from the atmosphere. Thus, knowledge of factors influencing forest cover can be used in the creation of policies capable of increasing carbon sequestration, and contributing to the goals and aims of the world in combating climate change. The study form can help formulate intervention strategies for climate change impacts. This implies the identification of those sectors that are at risk of being threatened by deforestation due to climate change and which we can avoid or minimize. The results can therefore be employed by the policymakers as a source of resourceful and utilitarian environmental policies. This comprises laws designed to safeguard sensitive forest areas, encourage sustainable land use, and lessen deforestation. To maximize the impact of conservation efforts, governments, and organizations can spend resources more effectively by concentrating their efforts on the regions most impacted by climate change. It is useful for farmers and agricultural planners to comprehend how shifting climate variables impact forest regions. Planning sustainable farming methods that lessen their negative effects on neighboring forests and lower the chance of deforestation can benefit from this. Understanding how climate and forest cover interact might help land use planners better balance the demands of both agricultural growth and forest preservation.

5.3 Ethical Aspects

We do not violate anyone's right to privacy or human dignity in the course of our working procedures or methods. Multiple free internet sites provided us with our data. Because of this, the data we used cannot be used to identify or harm any individual. We have not conducted any dangerous research on people or animals during this examination. Even still, the internet is where we obtain our information. When completing our project, we did not claim the intellectual property of any other company or individual. While we were working, we utilized our personal computers. Nobody else's tools or data have been utilized or pilfered by us. All of the following were upheld during our investigation: integrity, legality, honesty, and transparency.

5.4 Sustainability Plan

Several crucial actions must be taken to guarantee the long-term viability of research on how climate change affects forest regions. To make data accessible to researchers and policymakers, it is first imperative that it be regularly collected, stored, and made available through collaborations with meteorological and forestry institutions. Ongoing research requires securing long-term funding through grants and partnerships, encouraging interdisciplinary collaborations, and regularly upgrading models and approaches. With the help of environmental groups' activism, policy proposals based on the findings should be created and promoted to impact environmental legislation. To engage the community and promote a sustainable culture, public awareness campaigns, and educational initiatives are essential. To reduce their influence on the environment, research operations should implement green technologies and eco-friendly procedures. The effectiveness of the strategy will be regularly monitored and evaluated, and progress will be periodically assessed through impact assessments. Developing solid alliances with local people, non-governmental organizations, and the government will help to promote sustainable forest management. Last but not least, ensuring the long-term viability of forest resources will require creating plans to increase forests' resistance to climate change and control related hazards. This all-encompassing strategy aids in promoting sustainable development, reducing climate change, and protecting forests.

CHAPTER 6
CONCLUSION AND FUTURE WORK

6.1 Summary of the study

The main goals of the study are to determine how the annual changes in forest area and climate variability vary. Historical data was gathered for the study from a number of internet sites. The data trends were visually and quantitatively examined through the use of exploratory data analysis. The model was trained using four machine learning algorithms: ridge regression, gradient boosting, lasso regression, and random forest. We needed three months for that. I have to take a number of actions one by one in order to finish this research. The following is listed after the description of the entire work period:

Step 1: Study various research articles to understand the story.
Step 2: Collect data from numerous sources.
Step 3: data preprocessing.
Step 4: Exploratory Data Analysis
Step 5: Model Selection
Step 6: Use ridge regression, gradient boosting, lasso regression, and random forest as models.
Step 7: Model Evaluation.
Step 8: Prediction.
Step 9: UI Implementation.
Step 10: Verify the output or outcomes. The proper sequence of events was followed to complete our job. These models will facilitate further experimentation on other projects

The work was finished by completing these procedures in the sorted position. The use of this model for climate change and deforestation analysis will aid in the development of future initiatives.

6.2 Conclusions

These findings propose that suitable modeling strategies are embraced in the research. To predict the temperature from meteorological data. A comprehensive study showed that Linear Models which are Ridge Regression, and LASSO Regression models were defeated by non-linear models, particularly by Random Forest and Gradient Boosting. Although non-linear models performance was also good though. The ensemble approaches revealed a rather high accuracy for the stream of data. precision, such that the R2 values are over 95%, which provides a rather accurate fit when compared to the original form. dependencies present in the data can be present in the correlations that exist between the parameters we want to convey. Thus, the applied interventions of these models can make a substantial difference in forest management undertakings by enhancing planning, forecasting, and other activities. about adaptation to climatic effects in forest areas. This study provides insightful information for the improvement of the accuracy of the predictions that were utilized practically to predict future temperatures and stresses the importance of the choice of the model in the analytical work with environmental data.

6.3 Implications for further study

The work that this is referring to is about creating a management tool with the help of computational models to predict temperature. These models are like smart tools that can recommend some action or the likely event that might happen to temperature in the future. Models are compared depending on which of them performs the best. making these predictions. As for me, the most accurate model is the one that will be determined after further investigation of the given problem and can be used to discover ways to become more efficient in making decisions regarding the protection of forests. For example, if the model informs us that perhaps, this or that area of the forest might diminish in size due to things like beauty climate change, or human activities, and temperature may increase due to that. They can come up with strategies on how to ensure that does not occur. We could have to go and plant more trees or had to develop policies to help them protect the forest and the increasing temperature trend. By using these accurate models, we can also plan new projects to grow more trees in areas that are being deforested and need it most. This is useful in ensuring that we only limit our resources to what would best fit the space we have and ensure that everything is running as it should. the least harm to the environment and most good for the environment.

6.4 Limitations

Our approach to predicting future temperatures using historical weather data and forest area records, while promising, has several limitations. The accuracy of our predictions depends heavily on the quality and completeness of the historical dataset, spanning from January 1975 to October 2023. Any gaps or inaccuracies in this data could affect the model's performance. Additionally, our model assumes that past trends will continue into the future, which may not account for unexpected changes in climate patterns, extreme weather events, or shifts in forest area due to deforestation or reforestation efforts. Moreover, the estimation of future rainfall and forest area values is based on historical trends, which might not accurately reflect future conditions. Finally, while the user interface simplifies the prediction process, it currently requires manual input of the year and month, which may not be ideal for more comprehensive, automated forecasting needs.

References:

1. Ministry of Foreign Affairs available at <<https://mofa.gov.bd/site/page/6dde350b-1ca6-4c69-becd-a3f12cf14ac1/Bangladesh--An-Introduction#:~:text=Bangladesh%20is%20Located%20in%20the,Mizoram%20(India)%20and%20Myanmar.>> last accessed on 24-05-2024 at 7.45 PM
2. FAO. 2020. *Global Forest Resources Assessment 2020 – Key findings*. Rome. available at <<https://doi.org/10.4060/ca8753en>> last accessed on 24-05-2024 at 8.09 PM
3. Our World in data <<https://ourworldindata.org/deforestation>> last accessed on 24-05-2024 at 8.09 PM
4. [IPCC] Intergovermental Panel on Climate Change. 2013. Contribution to the IPCC fifth assessment report Climate Change WG I, 2013: The physical science basis summary for policymakers. Stockholm, Sweden: IPCC working group I.

5. Smith TM, Smith RL. 2009. Elements of ecology. 7th ed. San Francisco: Pearson Benjamin Cummings, United State of America. Chapter 29, Global climate change; p. 622646.
6. Khaine, Inkyin, and Su Young Woo. "An overview of interrelationship between climate change and forests." *Forest science and technology* 11, no. 1 (2015): 11-18.
7. Buizer, Marleen, David Humphreys, and Wil de Jong. "Climate change and deforestation: The evolution of an intersecting policy domain." *Environmental Science & Policy* 35 (2014): 1-11.
8. Kaggale available at <<https://www.kaggle.com/datasets/yakinrubaiat/bangladesh-weather-dataset>> last accessed on 22/05/2024 at 4.25 PM
9. The World Bank available at <<https://data.worldbank.org/indicator/AG.LND.FRST.ZS?locations=BD>> last accessed on 23/05/2024 at 3.26 PM
10. FAO available at <<https://www.fao.org/4/AC921E/AC921E01.pdf>> ast accessed on 24/05/2024 at 10.49 PM
11. Mirzaei, Ardalan, Stephen R. Carter, Asad E. Patanwala, and Carl R. Schneider. "Missing data in surveys: Key concepts, approaches, and applications." *Research in Social and Administrative Pharmacy* 18, no. 2 (2022): 2308-2316.
12. Mendelsohn, R. (2014). The Impact of Climate Change on Agriculture in Asia. *Journal of Integrative Agriculture*, 13(4), 660-665. doi:10.1016/S2095-3119(13)60701-7
13. Malhi, G. S., Kaur, M., & Kaushik, P. (2021). Impact of Climate Change on Agriculture and Its Mitigation Strategies: A Review. *Sustainability*, 13(3), 1318. doi:10.3390/su13031318.
14. James, R., et al. (2017). Characterizing Half-a-Degree Difference: A Review of Methods for Identifying Regional Climate Responses to Global Warming Targets. *Wiley Interdisciplinary Reviews: Climate Change*, 8, e457.
15. Oliveira, G. C., Arruda, D. M., Filho, E. I. F., Veloso, G. V., Francelino, M. R., & Schaefer, C. E. G. R. (2021). Soil Predictors are Crucial for Modelling Vegetation Distribution and Its Responses to Climate Change. *Environmental Science and Policy*, 120, 123-134.
16. Lutz, A. F., Immerzeel, W. W., Gobiet, A., Pellicciotti, F., & Bierkens, M. F. P. (2013). Comparing Climate Change Signals in CMIP3 and CMIP5 Multi-Model Ensembles and Implications for Central Asian Glaciers. *Hydrology and Earth System Sciences*, 17(9), 3661-3677. doi:10.5194/hess-17-3661-2013.
17. Kundzewicz, Z. W., & Stakhiv, E. Z. (2010). Are Climate Models "Ready for Prime Time" in Water Resources Management Applications, or is More Research Needed? *Hydrological Sciences Journal*, 55(7), 1085-1089. doi:10.1080/02626667.2010.513211.
18. Meinshausen, M., et al. (2011). Emulating Coupled Atmosphere-Ocean and Carbon Cycle Models with a Simpler Model, MAGICC6—Part 1: Model Description and Calibration. *Atmospheric Chemistry and Physics*, 11, 1417-1456.

19. Mitchell, D., et al. (2017). Half a Degree Additional Warming, Prognosis and Projected Impacts (HAPPI): Background and Experimental Design. *Geoscientific Model Development*, 10, 571-583.
20. Moss, R., et al. (2010). The Next Generation of Scenarios for Climate Change Research and Assessment. *Nature*, 463, 747-756.
21. King, A., et al. (2017). Australian Climate Extremes at 1.5°C and 2°C of Global Warming. *Nature Climate Change*, 7, 412-416.
22. Sanderson, B., et al. (2015). A New Ensemble of GCM Simulations to Assess Avoided Impacts in a Climate Mitigation Scenario. *Climatic Change*, 132(3), 499-511.
23. Nelson, G. C., et al. (2014). Economic Impact of Climate Change on Global Agriculture: Evidence from a Meta-Analysis. *World Bank Research Observer*, 29(2), 187-207.
24. Wong, P. P., et al. (2014). Climate Change Adaptation in Coastal Areas: A Review of Strategies and Approaches. *Ocean & Coastal Management*, 94, 76-83.
25. Nakashima, D., Galloway McLean, K., Thulstrup, H. D., Ramos Castillo, A., & Rubis, J. (2012). Integrating Traditional Knowledge and Scientific Approaches in Climate Change Adaptation: A Review. *UNESCO Working Paper*, Paris: UNESCO.
26. Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W., & Courchamp, F. (2012). Climate Change Impacts on Biodiversity and Ecosystem Services: A Review of Research Findings and Policy Implications. *Climatic Change*, 112(1), 93-106.
27. Lesnikowski, A. C., Ford, J. D., Biesbroek, R., Berrang-Ford, L., & Heymann, S. J. (2016). Policy Frameworks for Climate Change Adaptation: A Review of Country Experiences and Emerging Trends. *Environmental Science & Policy*, 66, 242-259.
28. Revi, A., Satterthwaite, D. E., Aragón-Durand, F., Corfee-Morlot, J., Kiunsi, R. B. R., Pelling, M., Roberts, D. C., & Solecki, W. (2014). Climate Change Adaptation in Urban Areas: A Review of Strategies and Practices. *Progress in Planning*, 95, 1-66.
29. Vermeulen, S. J., Campbell, B. M., & Ingram, J. S. I. (2012). Addressing Food Security and Climate Change Through Innovative Adaptation Strategies. *Global Environmental Change*, 22(4), 810-820.
30. Schipper, E. L. F., & Pelling, M. (2006). Sustainable Development and Climate Resilience: Integrating Environmental and Social Goals. *Climate and Development*, 8(1), 14-25.
31. Mollah, Tanjinul Haque, Anannya Mazumder, and Takaaki Nihei. "Current scenario and challenges of agricultural production to future food security in Bangladesh." *Journal of Agricultural Extension and Rural Development* 11, no. 11 (2019): 208-224.
32. Tukey, John Wilder. *Exploratory data analysis*. Vol. 2. Reading, MA: Addison-wesley, 1977.
33. James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. *An introduction to statistical learning*. Vol. 112. New York: springer, 2013.
34. Weather Online available at <<https://www.weatheronline.co.uk/reports/climate/Bangladesh.htm#:~:text=There%20are%2

0three%20distinct%20seasons,most%20parts%20of%20the%20country.>> last accessed on 29/05/2024 at 1.24 PM.

35. Behrens, John T. "Principles and procedures of exploratory data analysis." *Psychological methods* 2, no. 2 (1997): 131.
36. Kemp, Sarah E., May Ng, Tracey Hollowood, and Joanne Hort. "Introduction to descriptive analysis." *Descriptive analysis in sensory evaluation* (2018): 1-39.
37. Hegde, Vinayak, and M. S. Pallavi. "Descriptive analytical approach to analyze the student performance by comparative study using Z score factor through R language." In *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pp. 1-4. IEEE, 2015.
38. Camizuli, Estelle, and Emmanuel John Carranza. "Exploratory data analysis (EDA)." *The encyclopedia of archaeological sciences* (2018): 1-7.
39. Waskom, Michael L. "Seaborn: statistical data visualization." *Journal of Open Source Software* 6, no. 60 (2021): 3021.
40. Brodlie, Ken, Rodolfo Allendes Osorio, and Adriano Lopes. "A review of uncertainty in data visualization." *Expanding the frontiers of visual analytics and visualization* (2012): 81-109.
41. Turing available at <<https://www.turing.com/kb/random-forest-algorithm>> last accessed on 7/.6/2024 at 1.42 pm
42. Medium available at <<https://medium.com/@hemashreekilari9/understanding-gradient-boosting-632939b98764>> accessed on 7/.6/2024 at 1.44 pm
43. LinkedIn available at <<https://www.linkedin.com/pulse/lasso-regression-clearly-explained-bhabani-shankear-basak/>> accessed on 7/.6/2024 at 1.48 pm
44. MQL5 available at<<https://www.mql5.com/en/articles/11735>> accessed on 7/6/2024 at 1.50pm

Plagiarism Report

Thesis Report

ORIGINALITY REPORT

16%	14%	8%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Daffodil International University Student Paper	2%
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	1%
3	dokumen.pub Internet Source	1%
4	www.ipcc.ch Internet Source	1%
5	edepot.wur.nl Internet Source	<1%
6	iopscience.iop.org Internet Source	<1%
7	www.grafiati.com Internet Source	<1%
8	www.tandfonline.com Internet Source	<1%
9	espace.etsmtl.ca Internet Source	<1%