Utilizing machine learning techniques to enhance the accuracy of weatherprediction in Bangladesh

Submitted By

Md. Towqibur Rahman Khan

Student Id: 203-35-676

Department of Software Engineering

Daffodil International University

Supervised By

Mr. Khalid Been Badruzzaman Biplob

Designation: Lecturer (Senior Scale)

Faculty of Science and Information Technology

Department of Software Engineering

Daffodil International University

This thesis report has been submitted in fulfilment of the requirements for the degree a **Bachelor of Science in Software Engineering**

DAFFODIL INTERNATIONAL UNIVERSITY

©Daffodil International University

Department of Software Engineering

Faculty of Science and Information Technology

Supervisor Approval Form

Fall 2025	B.Sc. In SWE	Campus: DSC

Student Name	Student ID
Md. Towqibur Rahman Khan	203-35-676

Project/Thesis Information	
Project/Thesis Title	**Utilizing machine learning techniques to enhance the accuracy of weather prediction in Bangladesh**
Type of work	Thesis

Supervisor information	
Supervisor Name	Mr. Khalid Been Badruzzaman Biplob
Supervisor Initial	KBB
Completed Credit till now	139
How many credits in this semester	0
Supervisor Consent	☑ Yes ☐ No

Supervisor Signature

©Daffodil International University

APPROVAL

This thesis titled on "**Utilizing machine learning techniques to enhance the accuracy of weather prediction in Bangladesh**", submitted by **Md.Towqibur Rahman Khan (ID: 203-35-676)** to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and approval as to its style and contents.

BOARD OF EXAMINERS

Chairman

Dr. Hasan Mahmud
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 1

A.H.M Shahariar Parvez
Associate Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 2

Tapushe Rabaya Toma
Assistant Professor
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

Internal Examiner 3

Khalid Been md. Badruzzaman Biplob
Lecturer (Senior Scale)
Department of Software Engineering
Faculty of Science and Information Technology
Daffodil International University

External Examiner

Dr. Md Sazzadur Rahman
Professor
Institute of Information technology
Jahangirnagar University, Bangladesh

©Daffodil International University

DECLARATION OF THESIS AND COPYRIGHT

Author's Full Name	:	Md. Towqibur Rahman Khan
Date of Birth	:	03-07-2001
Title	:	Utilizing machine learning techniques to enhance the accuracy of weather prediction in Bangladesh
Academic Session	:	Fall 2020(203)

I declare that this thesis is classified as:

☐	CONFIDENTIAL	(Contains confidential information under the Official Secret Act 1997)*
☐	RESTRICTED	(Contains restricted information as specified by the organization where research was done)*
☑	OPEN ACCESS	I agree that my project to be published as online open access (Full Text)

Iacknowledge that Daffodil International University reserves the following rights:

1. The Project is the Property of Daffodil International University.
2. The Library of Daffodil International University has the right to make copies of the Project for the purpose of research only.
3. The Library of Daffodil International University has the right to make copies of the Project for academic exchange.

Certified by:

Towqibur

(Student's Signature)

(Supervisor's Signature)

Student ID : 203-35-676
Date: 17-12-2025

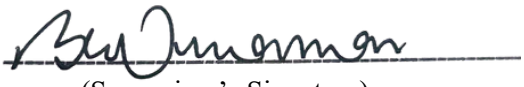
Mr. Khalid Been Badruzzaman Biplob
Name of Supervisor
Date: 17-12-2025

NOTE: * If the Project is CONFIDENTIAL or RESTRICTED, please attach a thesis declaration letter.

©Daffodil International University

SUPERVISOR'S DECLARATION

I hereby declare that I have checked this project and in my opinion, this project is adequate in terms of scope and quality for the award of the degree of Bachelor of Science.



(Supervisor's Signature)

Full Name : Mr.KhalidBeen Badruzzaman Biplob

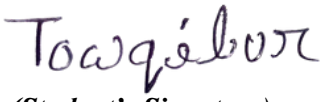
Position : Lecturer(Senior Scale)

Date : 17-12-2025

©Daffodil International University

STUDENT'S DECLARATION

I hereby declare that the work in this project is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Daffodil International University or any other institution.



(Student's Signature)

Full Name : Md. Towqibur Rahman Khan
ID Number : 203-35-676
Date : 17-12-2025

©Daffodil International University

Utilizing machine learning techniques to enhance the accuracy of weather prediction in Bangladesh

Submitted By

Md. Towqibur Rahman Khan

Student Id: 203-35-676

Department of Software Engineering

Daffodil International University

This thesis report has been submitted in fulfilment of the requirements for the degree a **Bachelor of Science in Software Engineering**

Department of Software Engineering

Daffodil International University

December 2025

©Daffodil International University

ACKNOWLEDGEMENT

In this walk of gratitude, my sincerest thanks go to Almighty Allah, whose endless compassion and strength enabled me to undertake this task with determination and resolve. Every step forward remained an act of Allah's grace.

I will always be thankful to express my sincere gratitude to my supervisor, Most. Mr. Khalid Been Badruzzaman Biplop, Lecturer(senior scale), Department of Software engineering at Daffodil international university without his continuous support, contribution and encouragement this would not have been achieved. This is not to say that her leadership and inspiration have not assisted me to achieve more than I had imagined in this work.

I would like to mention all these people that were so generous with their knowledge, their time, and support at different stages of this research. It is also necessary to state that these individuals all played an important role in the completion of this research in my life.

Also, I would personally appreciate to say a word of gratitude to all the faculty members of the Department of Software Engineering of Daffodil international university. I always found your constant support, academic help and belief in my abilities as motivational in my educational life.

It is not merely a milestone, but also a recognition of all the wisdom, patience and encouragement that many special people have given to my development. I am grateful.

©Daffodil International University

DEDICATION

This project was therefore done under the guidance of Mr. Khalid Been Badruzzaman Biplob, Lecturer (Senior Scale), Department of Software Engineering, Daffodil International University. I do not undervalue her advice, encouragement, and support in the course of this work development. I also confirm that this report is completely my own work and was never handed in wholly or part thereof to any institution or program to get academic credit or any other reason.

©Daffodil International University

ABSTRACT

The present days weather forecasting plays an important role in most industries, including the agricultural industry, transport, calamity management, and energy supply. Past forecasting procedures have been typically very effective; but the desired high standards of accuracy are largely not attained because of the inherent complexity and fluctuation of the weather. The present study will cover the use of machine learning methods to enhance the precision of the weather prediction in the various maximum and minimum temperatures of the weather of the next day. This project is more likely to offset the lower specs that most traditional models bring, with the aid of modern methods of computation, and through the latter, a more reliable foundation of decisions would have been established. It indicates that the preprocessing of data is not an option instead of a necessary step in the research since it involves cleaning the missing data, features normalization to maintain the integrity and consistency of the input data. The models of machine learning have been experimented and analyzed intensively in terms of significant performance metrics that may be indicative of the most suitable methods of temperature predictions. High techniques surely appear quite encouraging to the eliciting of those evanescent associations among variables towards much more precise and practical predictions. Also in addition to the technical input, the study would project into the future broader implications to society and the environment on the basis of enhanced weather prediction. Better predictions minimize the risk of calamities, enhance agricultural planning and aids sustainable management of resources such as water and energy consumption. The ecological concern is associated with the reduction of carbon footprint and the methods of successful incorporation of renewable energy sources. The other ethical concerns that it can face are equity in access, privacy of data, and responsible use of the predictive technologies. The article supports the possibility of machine learning to radically change the essence of one of the biggest problems confronting humanity as far as weather variability is concerned. In this way, the work preconditions the following studies on the borders that will combine an expanded set of data sources, advanced methods of modeling, and cost-effective deployment plans.

©Daffodil International University

Table of Contents

©Daffodil International University

©Daffodil International University

LIST OFFIGURES

©Daffodil International University

LIST OFTABLES

©Daffodil International University

Chapter 1

Introduction

1.1 Introduction

The field of weather prediction is a vital part of the contemporary society as it affects such aspects of the daily life as agriculture, disaster management, aviation, and daily life planning. In essence, these weather predictions offer the most suitable risk reduction principle through optimization of resources to protect the citizens. To take an example, proper identification of floods or hurricanes or heat waves will save many lives and save some economic loss. The weather forecasting has always been dominated by Numerical Weather Prediction models. These models invoke the assistance of the mathematical equations in simulating the behavior of the atmosphere on physical principles. Although these models to some degree are effective, there is a limitation attached to them. They tend to be computation-intensive and they can consume enormous computing capabilities and they usually can not implement the rapid variations of weather or localized phenomena. They remain dependent solely on the quality of initial conditions and data inputs and they are prone to errors in different circumstances.

This is achievable following the new advancements in data-driven methods of bettering weather forecasting. Machine-learning also offers methods of historical analysis of weather, discovering latent patterns in it, and prediction with no formal understanding of physical processes contained in it. These methods have now been capitalizing on the increased accessibility of various and massive data sets: weather data, sensor data, and satellite pictures [1].

Exploiting these information-based models increases effectiveness and precision in weather forecasts. Such a shift of focus towards less traditional, more purely data-driven strategies is new opportunities in the field of improvement of weather forecasts in a new direction that has not been available hitherto. The subsequent report evaluates the possible development of weather prediction by machine learning to address the conventional issues.

1.2 Research Motivation

Weather is a crucial factor that assists in the development of human activities and environment. Accurate weather forecasts are very crucial in most industries: agriculture, transportation, energy, catastrophe mitigation and community safety and security. The farmers have to rely on weather forecasts to schedule planting and harvesting seasons; the transportation network has to rely on weather forecasts to be sure of the mode safety and efficiency, and energy grids require extremely precise forecasts when it comes to using renewable sources of energy, such as solar and wind. It is also important to have reliable and timely forecasts in alleviating the impacts of very severe weather,such as hurricanes, flooding,and droughts, which have caused massive loss of life and economic damage.Despite the development in meteorology, setbacks are still experienced in the delivery of forecasts that are reliable at all times. Difficulties are compounded in rapidly varying weather or critical weather conditions, in which minor mistakes may be accompanied by serious consequences.

The majority of the techniques used in modern times to predict the weather are not able to capture the local weather features or predict better in the short term. Besides, locations where observational infrastructure is poor introduce additional complications in generation of forecasts because of poor information [2].

Due to climate change, the need to develop new and more efficient predicting techniques has grown even more pressing since the impacts of the extreme weather events are more intense and frequent. Better weather forecasts do not just reduce risks but also provide a basis of pro-active action in other ventures and societies. The chance to make a contribution to this significant sphere by trying to fill the gaps in the current forecast accuracy can be viewed as a good driving force behind the current study.

This article attempts to offer innovations in enhancing weather prediction, dependent on augmented data and innovations in data-driven techniques. Having the prospects of eliminating the existing constraints, such a study must result in more credible and informative weather projections in the best interest of the society and the environment [3].

1.3 Rationale of the Study

In fact, the issue of weather prediction has never been an easy task since not only are atmospheric systems extremely complex, but they are dynamic as well. The pattern of behavior of the weather, in its turn, is controlled by the plethora of interrelated factors, including the temperature, humidity, wind movements, and atmospheric pressure. Consequently, the therelies by itself is nonlinear behavior predisposition as stated in case of weather phenomena and is therefore hard to model. That, however, is infinitely more complicated in the case of the sudden changes, such as the storms, heat waves, and heavy rainfalls.

Traditionally, weather forecasting relies primarily on NWP-models, i.e. mathematical modeling of physical atmospheric processes. Although such models are very instrumental in development of meteorology, they have certain drawbacks. To start with, their dependence on high-quality initial data inevitably causes serious estimation mistakes in the outcomes of the forecast in case any inaccuracies or gaps are found in the observation data. Indeed, such models require an enormous amount of computation power and are inapplicable in real-time or predictions on local scale [4].

In other words, in extremely localized or even short-term weather conditions, the traditional tools have been lost, in fact, largely due to the fact that there is a sense of urgency in the present day, which requires fineness of data, capture of sudden changes. In fact, those gaps have manifested in accuracy either due to absence. real-time equipment in certain sites or during fast varying short term weather phenomena.It is these ventures that would be the keys to more accurate weather forecasts, timely warnings on any extreme weather occurrences and high-impact benefits to the society and economy. Novel methods of performing the weather forecast, more adaptive, efficient and accessible, are required, and capable of coping with intrinsic complexity in atmospheric systems, but taking into account constraints of the current practice. It discusses the solutions to such problems, by exploring partial solutions to the accuracy and reliability of the weather forecast, through the data-driven methodology.

1.4 Research Question

- More meteorological data sources would, therefore, enhance the weather predictions at different scales in regions and time.
- What is the fundamental flaw of the traditional system of predicting the weather and how would you suggest overcoming the weakness so that the precision of prediction can be improved?
- The integration of the local weather phenomena would be included in the manner of contributing the accuracy of the short-term forecast of the weather.
- What are variables that introduce errors in the current forecasting systems, how would they be reduced in dynamic and fast changes of weather?
- What is the role of enhanced weather prediction system contribution in proactive decisionmaking in the vital areas such as agriculture, transport and disaster management?

1.5 Expected Output and Objectives

The research is within the parameters of enhancing accuracy and reliability in weather forecasting- a rather important field considering a vast number of spheres of life like agriculture, transportation, disaster management, and public safety. Weather prediction does not merely remain the same as it can be affected by many variables. Among them, the common ones are temperature, atmospheric pressure, and pattern of wind. In that regard, this study has concentrated on the way the shortcoming of conventional forecasting methods can be experimented by other methods that penetrate deeper into the examination and interpretation of intricate trends in weather information.

- **The objectives of the study are:**

Accuracy of Forecasting: developing methods that would improve accuracy of forecast at localized and short scales that are not so good.

Smoothing Gaps in Data: Efficiency of use of available data, considering real treatments of incomplete or incoherent sets of data; this is a popular truth in meteorology.

Response to Multiple Scenarios: Construction of forecast models which perform well across a variety of conditions, such as extreme weather events and low density infrastructure of observation across the region. Societal Benefits: This study will provide knowledge and resources to reduce such impacts of unfavorable weather condition in preparing and making decisions of persons and citizens, firms, and governments.

The given research project helps decrease the gap between the increased requirements associated with the accuracy of weather forecasting and the lack of it that used to be previously present, thus paving the path to the improvement that must not only promote the science as such but also serve society in general.

1.6 Report Layout

The introduction, objectives and key research inquiries of the study are presented in Chapter 1, the introduction. Chapter 2 contains short synopses of the literature review. Chapter 3 explains the methodology suggested in detail. Chapter 4 presents the results of the experiment of the paper and discusses them. The fifth chapter addresses the sustainability plan, societal and environmental consequences and the ethical factors. The sixth chapter brings to an end the current research and develops a plan of actions in further activities.

Chapter 2

Background

2.1 Preliminaries/Terminologies

Weather forecasting has been a subject which has been researched upon in the past few decades as it affects practically every facet of human life and nature. More precise weather predictions are needed to reduce the deplorable impact of severe weather conditions, optimize the use of resources and make the population safe. Over several decades, different weather forecasting methods that employed different techniques to make suggestions have been advanced starting with the traditional numerical models to the modern sophisticated data-driven models [5].

The writings on weather forecasting have shown how complex and dynamic the atmospheric system is and hence it always poses challenges. Although conventional approaches towards forecasting which are founded upon mathematical modeling of atmospheric physics have achieved considerable milestones, nonlinear interactions and abrupt variations of weather patterns are still a limitation to the approaches. Local and real time requests of meteorological forecasts reveal a weakness in traditional methods to be computationally inefficient and imprecise.

Increase in the available data and capability of processing has enabled giving new directions to the process of weather prediction within the recent years. Other studies have examined different data sources such as satellite data, sensor networks and historical weather records to achieve better and dynamic forecasting techniques. These studies stress that something more creative is needed beyond the existing models' capabilities to face growing challenges brought forth by climate change and extreme weather events.

©Daffodil International University

2.2 Related Works

Holmstrom et al. proposed a model to forecast the maximum and minimum temperatures over the next seven days using data from the previoustwo days [10]. They developedtheir predictive model using linear regression and an extended version called the functional linear regression model. While these models produced some success, they did not meet the level of accuracy that professional weather forecasting services achieved for the seven-day forecast. The models were much better at longer-range forecasts or even determining trends in the weather after the seven-day forecast period.In another work, Krasnopolsky and Rabinovitz differently proposed a physical weather process-based model with the addition of neural networks [11]. By simulating the demanding atmospheric behavior, their work leveraged the computational strength of neural networks for enhancing weather prediction precision.When the weather forecasting problem is posed as a classification task, and SVM based technique has been used for predicting weather conditions in Radhika et al. used a different strategy [13]. This proved to be a novelusage of machinelearning for purposesof weather forecasting, particularly in the classifying and detecting different meteorological regimes.

A data mining-based predictive framework was presented in [14], whichaimed to find fluctuating patterns in historical weather data. This model employed the Hidden Markov Model for its prediction and integrated k-means clustering to extract meaningful observations from the historical data of weather conditions. The integration of HMM with clustering was used to approximate the future weather conditions based on past patterns. It provided a structured approach toward the analysis and forecasting of changes in weather.

Grover et al. proposed a hybrid approach which combines discriminatively trained predictive model sand deep neural Models for Forecasting Weather [9]. The model in particular concentrated on the statistical dependency relationships between a sequence of weather statistics so that temporal evolution aspects could be modeled in detail. When the prediction technologies were fused, their model also faired best atpredicting during complex weather conditions.

Another innovative approach was proposed by Montori et al. The authors have designed a crowdsensing-based solution, which serves to offer environmental phenomena monitoring due to voluntary exchange of the data captured via smartphones of individual users [12]. In this regard, they developed an architecture known as SenSquare, which is meant to aggregate IoT source data and crowdsensing Web. The obtained data was harmonized in an accessible form, which could be used in such applications as smart city environmental monitoring. The application of crowd sensed information presented the new paradigm in weather monitoring and forecasting because of the character of real-time environmental data distributed by the participants.

These were different strategies, whose strategies were not similar to the ones which incorporated machine learning models to hybrid systems, but instead to crowdsourced data, yet not a single piece of evidence suggested linking data of the adjacent geographical locations to enhance the accuracy of weather forecasting. This gap in this respect gives the future research the room to examine how spatial information in neighboring areas would be exploited to enhance the operation of the forecasting.

2.3 The Problem's Scope

The complexity of weather forecasting is also present in the very dynamics of meteorology phenomena: they are nonlinear, quite fast-changing, and extremely unpredictable. Areas of relation that cannot be easily observed through more conventional methods of statistics over longer time-frames when used to analyse historical data.. While professional forecasting servicesare extremely good at short-term forecasts, their skills deteriorate once the forecast length goes beyond a few days. This shows that long-term weather prediction contains more uncertainties to be dealt with through more robust methodologies.

Despite recent breakthroughs in machine learning for weather forecasting, many of these approaches operate on a single isolated dataset and do not considers patial dependencies. Most of the weather patterns are indeed interconnected with other regions since the general set-up of atmospheric systems is one which is interlinked. Any model not considering this spatial context runs the risk of providing limited or incorrect predictions. This gap, if filled by adding data from surrounding regions, can bring a lot of improvement in the field of precision and credibility regarding forecasts of weather, especially those of regional and localized phenomena [6].

Hybrid solutions in which physics based models are integrated with machine learning techniques have been able to attain some potentiality of filling this gap. As an example, the neural networks are suitable to model nonlinear dynamics that emerge as a result of weather systems. Nonetheless, the majority of them are based on history and do not take into account the real time information about neighboring regions. The disaggregation of such data in this fashion impairs seriously the adaptability of such data in the event of a sudden change in weather-like storms or change in the temperature which is truly essential in effective forecasting.

Whereas methods like data mining and probabilistic models like the Hidden Markov Models have been used in the identification of recurrent patterns with the case of past weather, majority of methods that are developed are usually deficient in real-time dynamic scenarios. Likewise other clustering algorithms, like the k-means, are pre-disposed towards identifying trends as opposed to isolating the temporal and spatial complexities that are apparent in predictions of the weather. The curtails are therefore an argument in favor of the development of those models that will be able to combine historical data information with real-time and spatially informed feeds [7].

These are some of the challenges that could be overcome with the assistance of the new technologies in crowdsensing and the environmental monitoring systems based on IoT. These systems can collect real-time data at a collection of devices and consumers and give a stream of rich, localized weather data. The existing practical implementations of such technologies are more on an environmental context than actual weather surveillance. A combination of this with predictive modeling particularly with the data of other nearby regions can provide a direction towards yet even better and more enhanced forecasting systems.

2.4 Challenge

The issue of weather prediction is associated with several systems of challenges since the atmospheric system is a dynamic system and the interaction between different parts of the system is extremely complex. The nonlinear association between different meteorological variables, whereby temperature, humidity, wind velocity, and pressure of the atmosphere are part of the variables, is one of the key problems as they are also dependent variables on external elements, including the sunshine and topography.

Once more there is the problem of the availability and quality of data. Weather forecasting is a reliable process that needs extensive data in high resolution obtained a given duration of time by different sources.

Unluckily, the noise, the absence of data and even the gaps in observation may decrease the precision of predictions made by the model. The power of the models to predict accurately in such spots to finer resolutions, is limited by the difficulty in obtaining real-time data of remote or inaccessible locations.

The other issue is the inclusion of spatial dependency in the weather forecasting models. Atmospheric phenomena are traditionally interrelated in a spatial collection of conditions in one place can condition the patterns of other places. Most of the existing models have the general focus on isolated data which fail to make use of this spatial information resulting in not-so-accurate regional forecasting. It involves sophisticated methods and substantially greater calculation capabilities to combine information about adjacent places making it difficult to handle with large scale situations [8].

Other significant challenges are the scaleability and computational complexity of the models. The majority of existing machine learning systems especially deep learning models are intensive in terms of memory and computation requirements both when training and in deployment. As an example, models that combine physical-based simulation and machine learning algorithms can involve high-intensive calculation; hence, real-time prediction is barely possible to yield. The question of balancing model accuracy and computer computations is one of these challenges that have been persistent.

Chapter 3

RESEARCH METHODOLOGY

3.1 Proposed Methodology

The research design to be used will be based on the development of an effective model of prediction of the maximum and minimum air temperatures of the next day through the use of machine learning algorithms. Massive processing of missing values in form of imputation of missing values in the data will also be undertaken using the K-Nearest Neighbors (KNN) Imputer just to maintain coherency of the data without losing the connection between the variables.

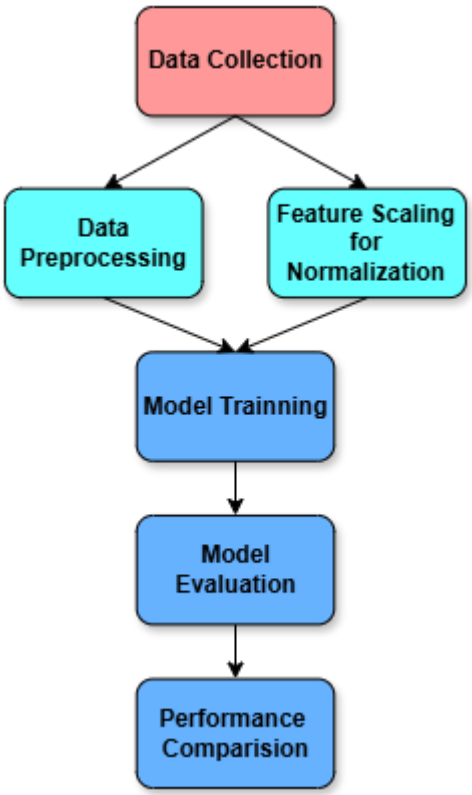


Figure 3.1: Proposed Research Methodology

This dataset contains observation and forecast information about the weather from the years 2013 to 2017. Min-Max Scaling normalization was applied to the dataset. During the normalization process, the features were scaled to a standard range of [0, 1]; hence, it avoided biases that might result from the difference in

©Daffodil International University

magnitude among temperature, wind speed, and solar radiation. After the preparation of data, various machine learning models were trained on a multi-output regression approach so as to make concurrent predictionsof maximum and minimum temperatures. The model development included a number of trial models that vary from simple linear regression to decision trees, random forests, and gradient- boosted methodssuch as XGBoost, LightGBM, and CatBoost. Some of the performance metricsused in the analysisinclude the R^2 score, mean squared error, root mean squared error, and mean absolute error. The methodology utilized in this approachwas such that through multi-output regression, it was ableto identify interdependence between target variableswith the expressaim of optimizing the prediction accuracy of the variables. This approach was step-by-step to ensure the models were data- driven with characteristics peculiar to the weather forecast.

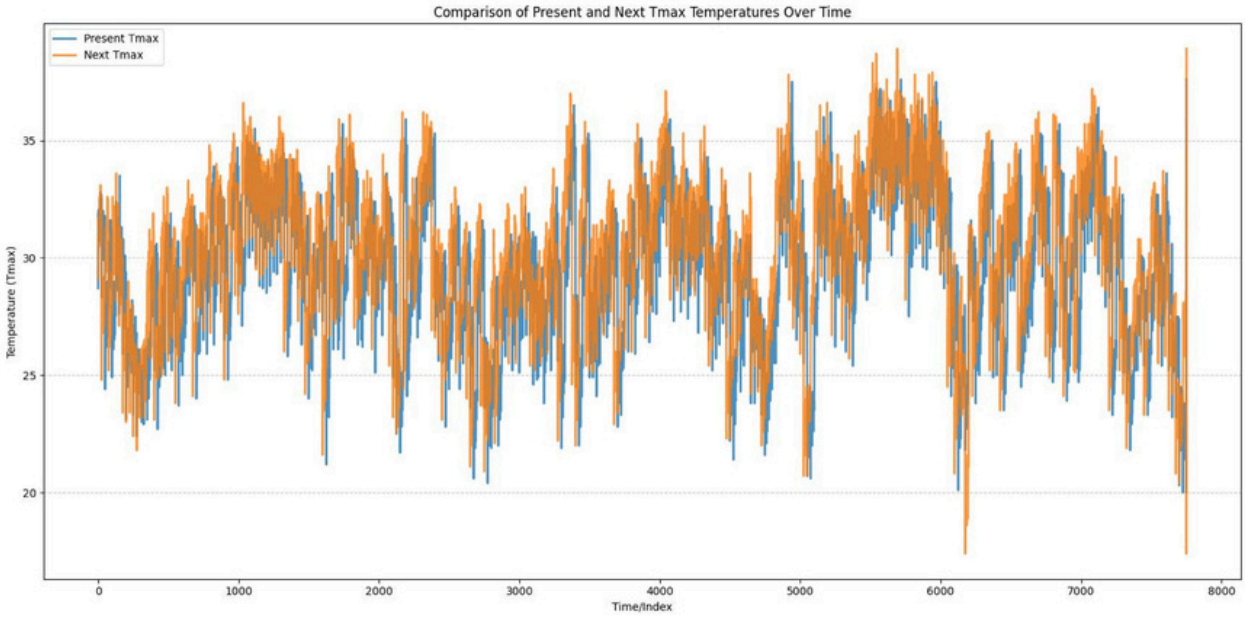


Figure 3.2: Present_Tmax vs Next Tmax

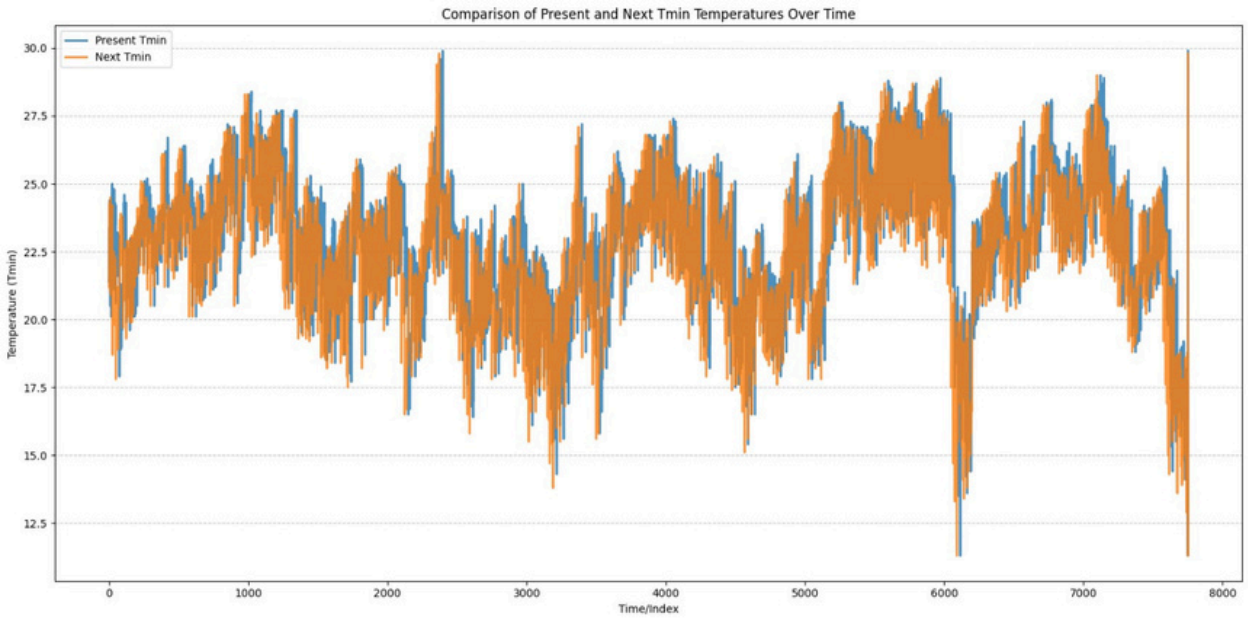


Figure 3.3: Tmin vs Next Tmin

3.2 Data Cleaning

For any data-driven project that works with real-world data, such as weather forecast data, data cleaning is necessary. The performance and reliability of the machine learning models may be severely affected by missing values in some features of the dataset being employed for this study. The missing values in this study were treated as follows:

i. Identification of Missing Values

- Upon the initial load of the dataset, a detailed examination for any missing values was performed. The. isna(). The sum() function was applied on which we could it the number of null values by column.
- Features such as station and date were dropped, because they had been identifiers rather than features and did not help in predictive modeling.

ii. Imputation Technique

KNN imputer is a robust way of handling missing values by imputing them on the basis of the similarity of data points. The KNN imputer uses the concept of neighboring data points to predict the missing values, considering only those imputed values that are in tune with the distribution of the existing data.

- The KNN imputer was initialized with 2 neighbors to balance accuracy and computational efficiency.
- The imputation process was done on all the numerical columns, excluding the target variables, Next_Tmax and Next_Tmin, so that the integrity of the predictive modeling was preserved.

iii. Benefits of Using KNN Imputation

- It preserves the relationships between features.
- Unlike mean or median imputation, KNN considers the local structure of the data, thu yielding more accurate imputations.
- It prevents bias that could occur because of constant imputation strategies.

3.3 Feature Scaling for Normalization

Inorder to ensure that each feature is contributing the same towards prediction,feature scaling becomes a necessary pre-processing step. Weather prediction data include measurements with very different ranges, such as the temperature (in ∘ C), wind speed (m/s) and sun radiation (Wh/m²). Unscaled, models can make grossly inaccurate predictions by assigning large weights to features with larger magnitudes. **Why Min-Max Scaling?** Inorder to ensure that each feature is contributing the same towards prediction,feature scaling becomes a necessary pre-processing step. Weather prediction data include measurements with very different ranges, such as the temperature (in ∘ C), wind speed (m/s) and sun radiation (Wh/m²). Unscaled, models can make grossly inaccurate predictions by assigning large weights to features with larger magnitudes.

Min-Max Scaling: Why Use It?

Min-Max Scaling was applied as the scaling method for this project. It is ideal for distance based methods (such as regression or decision trees) because when it scales the data, it represents the data in a fixed range of [0, 1].

The Min-Max Scaling formula is:

$$X_{\text{scaled}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \quad \text{.....(1)}$$

Where:

- X: Value of the original feature NdrFc
- Xmin: The smallest value of the feature • Xmax: The greatest value of the feature
- Xscaled: A scaled 0–1 value.

Steps in Feature Scaling

The sklearn. preprocessing module was created and all features on train set were transformed by that. Non-numeric columns, such as station and date, were removed from the dataframe since they were nnumeric and shouldn't be scaled. To ensure that the input features and output variables (Next_Tmax, Next_Tmin) are compatible with machine learning model the scaling was performed independently.

Advantages of Scaling

So the wide range features will not dominate the small range features, and all on the same scale. Acceleration of ConvergenceVarious optimization algorithms that are based on the gradient force. Enhances performance of machine learning models, particularly those sensitive to magnitude of features (e.g., SVMs, KNN-based classifiers, and neural networks).

3.4 Multi-Output Regression for Simultaneous Prediction

To handle two tasks (one-time and future) or more concurrently, we propose to use multiple-output regression analysis. To simulate both Next_Tmax and Next_Tmin simultaneously, a multi-output regression model was also considered. The supervised learning approach of multi-output regression predicts multiple target variables based on one set of input features. Because of the intimate association and similar prediction properties between Next_Tmax and Next_Tmin, such a approach can be used for weather forecasting.

Why Multi-Output Regression?

- **Related Findings:** Shared contributing factors such as low humidity, sun radiation, and wind velocityresult in a natural association between maximum and minimum temperature.

- **Efficiency** A one liner multi-output model does it all at once rather than creating a separate model for each dependent variable at a time.

Single-output regressors were transformed into multi-output regressors using the class
Multi Outpu tRegressor of the sklearn. multioutput library. Several base regressors were considered as the underlying models:

- **Linear Regression:** To establish whether the model can perform well, apply the linear regression.
- **Ridge Regression:** For preventing possible over-fitting by adding the regularization.
- **Decision Tree Regressor:** For shape of your curve, you need to go for decision tree
- **Random Forest Regressor:** The Random Forest Regressor algorithm is an ensemble learning approach for improved accuracy.
- **XGBoost:** XGBoost was also included to enhance the performance using gradient boosting.
- **LightGBM:** Because of its accuracy and fast computation time.
- **CatBoost:** Toachieve better performance and process the category features.

Model Training and Evaluation

i. **Training:** Each model was trained by Multi Output Regressor on the scaled train dataset. werede finedas Next_Tmax and Next_Tmin, and inputs as the scaled features Outputs

ii. **Prediction —**Eachfor two target variables were generated for test.Success.

iii **Evaluation Metrics:**

- **R2 Score:** It tellsus the just of the variance in the dependent variable can be explained by independent variables.
- It is a sole average of the amount that our estimates were off targetdenoted as Squared Erroror simply (SE)and therefore it should be an 22, Mean SQuare error or the \(MSE\). The square root One ratio that is easy to interpret in the same unit as the target variable is provided

by the Mean Squared Error (RMSE).

- The average magnitude of the errors is quantified by the Mean Absolute Error (MAE).

Multi-output regression, in order to improve the prediction accuracy, leverages the correlation of multiple target variables. trains a single model instead of two, which avoids duplication.

3.5 Algorithm Description

Linear Regression : Linear regression is the baseline model for many machine learning tasks. It works by constructing a linear equation to explain the relationship between the independent variable(features) and dependent variable(target):

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon \quad \text{.....(2)}$$

Here

- Y is the response variable.
- X 1, X 2,…, X n are the predictors at different levels (covariates). • β 0, β 1,…, β n, β are coefficients or parameter estimates of these covariates and error term is denoted by λ.

Interpretable and computationally fast, linear regression assumes that predictors and targets are linearly related. Its simplicity, on the other hand, may work to its detriment, especially when recording complex non-linear relationship. However, it is an important starting point for evaluating the baseline performance of our data.

Ridge Regression

Ridge regression improves upon linear regression by including a regularization term in the cost function to prevent overfitting:

$$\text{Cost Function} = \text{RSS} + \lambda \sum_{j=1}^{n} \beta_j^2 \quad \text{.....(3)}$$

(coefficients are β j, regularization parameter is λ). The Ridge regression enhances the generalization to new data and reduces complexity of a model by penalizing coefficients those are large. It is especially useful when you have multicollinearity in your dataset, as it reduces variance of the estimates. Ridge regression is thus a powerful alternative to linear regression particularly for work with highly correlated or noisy data features in weather prediction.

Decision Tree Regressor

The Decision Tree Regressor is a non-parametric method, meaning that it does not use any pre-assumption about the form of solution of the model and divides data into subsets recursively by feature values to fit them. To predict continuous targets, the tree structure partitions data at nodes using threshold values that minimize a loss function—for example, mean squared error. Decision trees are attractive for recording super-linear relationships and feature interactions. Interpretable The second biggest strength of their rules lays in the fact that they are interpretable – it is easy to observe and understand the "rules" for a specific forecast. But in this study, ensemble methods were useful to prevent overfitting of which the decision trees are susceptible, especially with deep tree.

Random Forest Regressor

To overcome overfitting and enhance precision, the Random Forest ensemble learning method is a combination of predictions generated by multiple decision trees. Each tree is trained on a bootstrap sample of the training set, one based not all or none but some those are also evaluated here with average for regression. Random Forests are also suitable for large datasets with screenful tools, and can avoid overfitting due to the ensemble averaging effect. It also provides feature relevance scores, which can help you understand how each piece of information is impacting the model's predictions. This is wellsuited for weather prediction, as it can handle non-linear data follow distributions.

XGBoost (Extreme Gradient Boosting)

A series of decision trees is built in succession using the powerful gradient boosting model XGBoost, where each tree attempts to correct its predecessor's mistakes. Its success is thanks to the fact that XGBoost can boost a custom objective function using second-order gradients (the Hessian) for more accurate and efficient boosting. It also provides distributed computing to speed up training on large data sets and includes regularization parameters to prevent overfit. XGBoost is beloved for its ability to scale and adapt, both of which are necesary when you're wrangling a messy dataset like hourly weather forecast data (because weather features are complicated and non-linea).

LightGBM (Light Gradient Boosting Machine)

Another highly efficient gradient boosting algorithm is LightGBM. Different from other boosting algorithms, LightGBM finds the best split positions on a histogram (used to achieve more fine-grained splitting), rather than using a pre-sorted partition for filtering and histogram construction in conventional decision tree learning. This effectively reduces training time and memory overhead of finding optimal split points since the search space is much smaller. In addition, it applies techniques such as leaf-wise tree growth, which is better in error reduction by nature than level-wise one. As previously shown in weather forecasting, LightGBM can perform exceptionally with high-dimensional or large-dimension datasets. Its applicability to practical problems is also improved with its support for nominal attributes and sparse input.

CatBoost

CatBoost does something similar to the gradient boosting method and is able to deal with categorical data without much pre-processing (like one-hot encoding). It reduces overfitting and achieves high accuracy using techniques such as oblivious trees and ordered boosting. Ordered boosting remains robust by avoiding the model to learn from future during training. On mixed-type datasets gradient boosting with CatBoost is often better than others because of its high speed. The capability of CatBoost to model complex relationships between meteorological parameters and its computational efficiency was particularly useful in this study.

Chapter 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Overview of Model Evaluation Metrics

This section summarizes all the evaluation metrics employed in assessing predictive model performance. The chosen metrics include R^2 Score, Mean Squared Error-MSE, Root Mean Squared Error-RMSE, and Mean Absolute Error-MAE. These metrics are put into work to comprehensively evaluate the models in predicting next-day maximum Next_Tmax and minimum temperatures Next_Tmin. R^2 Score refers to how well the model captures the variance in the target variables. A higher value toward 1 indicates better performance. MSE, RMSE, and MAE offer error measures in different scales and thus enable both statistical and practical insight from the models' predictions. It now prepares the way to proceed to the ends of comparing the individual model performance and their own strengths and weak points.

4.2 Experimental Results & Analysis

Model	R2 Score	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)
Linear Regression	0.840266	1.843438	1.357733	0.996920
Ridge	0.840263	1.843230	1.357656	0.996871
LightGBM	0.796422	1.921213	1.386078	0.998687
RandomForest Regressor	0.770867	2.121625	1.456580	1.070038
CatBoost	0.770075	2.013041	1.418817	1.018908
XGBoost	0.743892	2.265336	1.505103	1.090621
DecisionTree Regressor	0.692178	3.470552	1.862942	1.397839

Table 4.1: Model Evaluation Table

The evaluation results of the two machine learning models for the next day's maximum and minimum temperatures showed that each algorithm presented different strengths and weaknesses. To compare and evaluate their accuracy and predictive capability, the metrics of R^2 Score, Mean Absolute Error (MAE), Mean Squared Error (MSE) and Root Mean Square Error (RMSE) were considered. Linear Regression and Ridge Regression yielded the best overall performances, with R2 Scores of 0.8403 and 0.8402, respectively. The ability of the two models to explain linear relationships between predictors and outcomes was indicated by R2 values of about 84% for all target measures. They were probably very stable at producing predictions all over the place, as their MSE, RMSE and MAE values had the lowest results among all other models. The small/garbage collector effect of regularization in the Ridge Regression for this dataset is confirmed by these sort of difference between both models.

LightGBM was the best-performing ensemble method with an R2 Score of 0.7964. Its competitive RMSE and MAE of 1.386 and 0.999 show that it could model a non-linear pattern in the data quite well, although it is less accurate than linear forms. It was superior to other ensemble approaches such as Random Forest and CatBoost because of its histogram-based approach and efficient leaf-wise tree expansion. Even though LightGBM is a non-linear model and does an awesome job modeling the data, some of the issues with the dataset inability to capture those peaks I discussed above are reflected in its slightly higher error numbers vs. GLMs.

Both Random Forest and CatBoost performed almost equally well with the R^2 Scores of 0.7709 and 0.7701. Even Random Forest did a better job as it reduced overfitting and the variance. This allowed it to achieve an RMSE of 1.456and an MAE of 1.07.CatBoost utilized the gradient boosting method and its RMSE was 1.419, MAE was 1.019 which was slightly better than the Random Forest. Both models performed

competitively in this competition but they couldn't match LightGBM's overall performance.

The other ensembles had a worse score by XGBoost, as it was 0.7439 (R^2 Score). Where RMSE is 1.505 and MAE is 1.091, demonstrates that Gradient boosted trees could not handle this dataset as well as LightGBM and CatBoost did. The findings of this paper reveal that XGBoost's defaults may not work best for weather prediction (although it works well generally across the board very popular choice).

The Decision Tree Regressor performed the worst of the different models that were used, with a performance of an R^2 Score of 0.6922.Its very high MSE of 3.47 and RMSE of 1.863 emphasizeits inability to generalize well, more than likely due to overfitting on the training set. While it gave insight into feature importance, by itself, this fell below the mark necessary for accurate weather prediction.

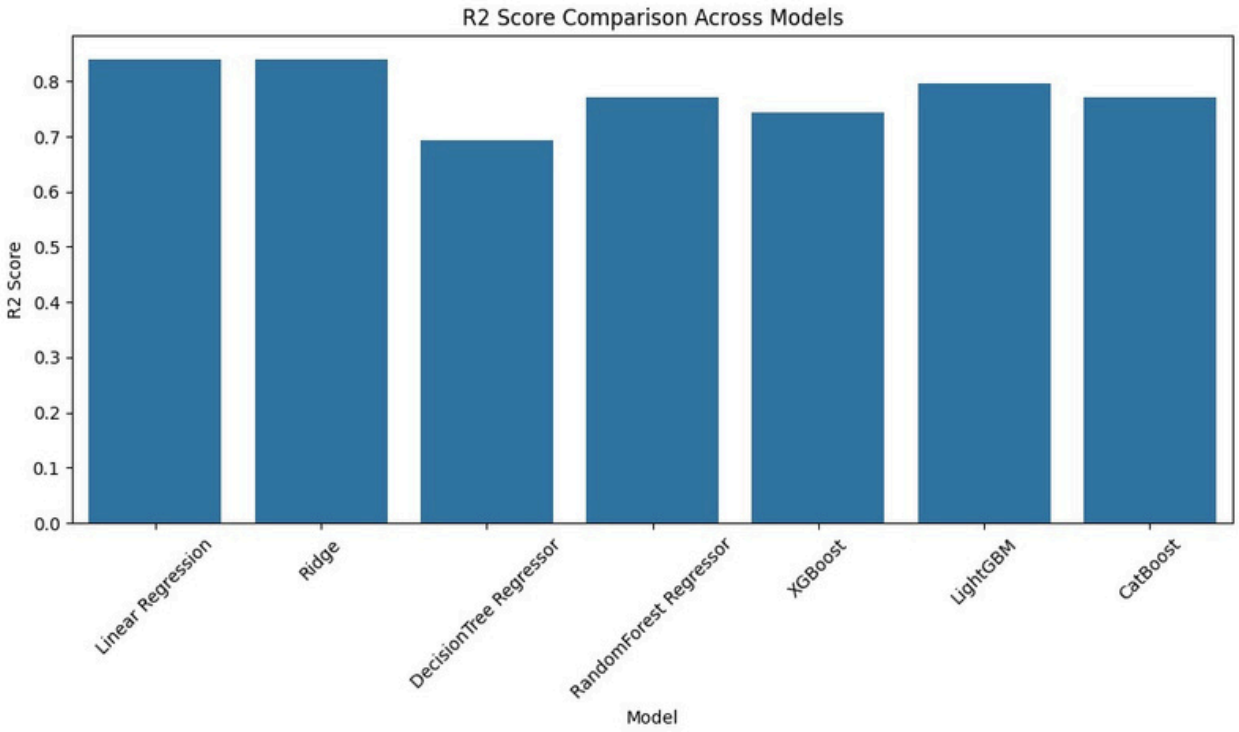


Figure 4.1: R2 Score Comparison Across Models

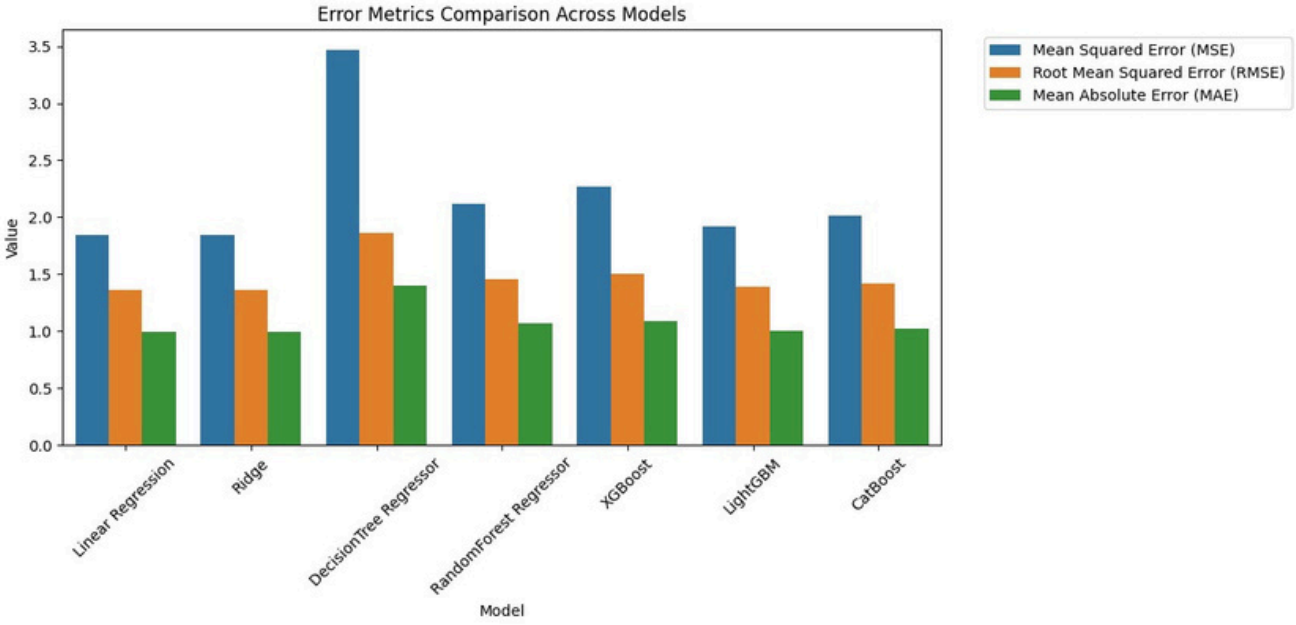


Figure 4.2: Error Metrics Comparison Across Models

4.3 Performance Analysis of Baseline Models

The baseline models, Linear Regression and Ridge Regression, gave a solid ground to understand the predictive capability of the dataset. Both achieved an R^2 Score of approximately 0.84, reflecting each model's capability in explaining 84% of the variance in the target variables. The small difference between their metrics gives a value of 1.8434 for the MSE of Linear Regression against 1.8432 for Ridge. This marginal difference indicates the minimum effect brought in by Ridge Regression due to its regularization. These results reflect the fact that, though simple, linear models are pretty powerful when there is a high degree of linearity between features present within the datasets. However, linear models have their shortcomings in capturing nonlinear interactions; hence, further exploration of more complex algorithms is warranted in sections to follow.

4.4 Comparative Analysis of Ensemble Learning Models

We also examined some ensemble learning models; Random Forest Regressor, LightGBM, XGBoost and CatBoost are capable of identifying complex nonlinear information in our dataset. The model that perform the best on this dataset is LightGBM with R^2 Score of 0.796 followed by Random Forest(0.771), CatBoost(0.770) and XGBoost(0.744). The advantage of the efficiency and accuracy in LightGBM may come from its histogram-based and leaf-wise processing algorithm. Random Forest worked well since it took an average of many decision trees, which helped cut down overfitting and variance. CatBoost did perform very well, especially with categorical features but not as great as LightGBM. The XGBoost and other gradient-boosting models didn't work either. These results indicate that ensemble methods add a great amount of accuracy in prediction as compared to linear models but with the cost of computational time.

4.5 Insights from Decision Tree Regressor

This makes the Decision Tree Regressor a non-linear model that can study the interactions between features of the dataset. Being of an R^2 Score of 0.692, it is poorly performing with respect to the ensemble models. Its high MSE of 3.47 and high RMSE of 1.86 show how susceptible it is to overfitting since single trees usually generalize poorly on unseen data. Although Decision Tree Regressor performed worse, it also contributed a lot of value in feature importance that informed the design and optimization of the ensemble models. This section discusses the trade-offs between interpretability and performance inherent in single-tree models compared with ensemble methods.

4.6 Error Analysis and Interpretation

Once you have places to improve, you need to understand why each model failed. Ridge and Linear Regression also performed well in the global error metrics, however they did not perform especially good for non-linear patterns, specially on the season when Next_Tmax and Next_Tmin are at extremes. While they sometimes made larger errors for moderate values, at the extremities in particular ensemble models (and LightGBM and CatBoost especially) were generally accurate on these types of data, possibly indicating some type of overfitting to this data. The implications of the MAE values—which fluctuated between 0.996

for Ridge and 1.39 for Decision Tree Regressor—are also presented in this section, underlining the practical applicability of these shortcomings in real-world weather prediction.

4.7 Discussion

The model performance analysis is of particular interest to know how the machine learning models are effective at predicting both maximum and minimum air temperatures for the next day. Linear and Ridge Regression With the top R2 scores of 0.8403 and 0.8402, Linear (Ridge) Regression differentiated itself with excellent ability to exploit linear patterns to explain variance in the data. The j48 and REP trees also exhibit their reliability as baseline models, which can be seen from the low error-values (RMSE/MAE). However, due to their simplicity, these methods are not very good for modelling weather extremes when wind is involved, as they may fail to model non-linear properties of turbulence. The dataset may not have high multicollinearity or overfitting issue as Ridge Regression, able to constrain the weights, only slightly better than Linear Regression.

LightGBM, with a R^2 Score of 0.7964 is the best ensemble model which suggests that it was capable of capturing high-dimensional data points and non-linear relations. It outperformed other ensemble methods such as Random Forest and CatBoost in accuracy and efficiency, demonstrating its suitability for real applications. Those libraries along with Random Forest and CatBoost turned out to be similarly performing; their error metrics were just a little worse, so perhaps they miss some tuning for this purpose. Although XGBoost is famous for its robustness, it's ranking was low in this study with an R2 Score of 0.7439 possibly due to less-than-optimal default parameters. Finally, in terms of generalization performance, Decision Tree Regressor also had the lowest R2 score of 0.6922 which demonstrated the limitation as an individual model for hard prediction tasks. These findings illuminate the compromise between interpretability, computational efficiency and predictive accuracy, and, in our study LightGBM proved to be the most suitable compromise for weather prediction.

Chapter 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

Surely there are enormous social benefits from incorporating machine learning into weather prediction. Being able to predict the weather accurately is critical for saving human lives, property and livelihoods. This research will assist in the preparation and planning for, as well as resilience actions against, the impacts of climate variability by improving forecasts for extreme weather events (as well normal variation) such as heatwaves, cold spells or rapid temperature changes. Precautions, evacuations and dialogue of resources- for mitigating the adverse effect of weather related hazards can be effectively planned by government agencies/ disastermanagement groups well in advancethan being affectedas a result of earlyand accurate forecasts generated out by meteorological department.

Secondly, weather sensitive and/or dependent industries such as energy production, transport and agricultural systems would greatly benefit from more accurate weather prediction. Accurate temperature forecasts can also assist farmers in determining precisely when to plant, irrigate and apply pesticides, to maximize crops and minimize losses. Prefix / At the same time, precise weather and temperature predictions would provide transport sectors such as shipping and aviation with equally beneficial gains in operational efficiency and safety. On the energy front, the predictions would help in improving the generation management of wind and solar power suppliers, which would alleviate the nonrenewable demands. In this perspective, the externalities of the research on society are not merely safety oriented, but also in terms of quality amelioration of life, effectiveness and long run economic safety of resources

5.2 Impact on Environment

The new machine learning weather forecasting resulting in proactive environmental management will certainly have a great impact on the environment in a positive way. Weather forecasting allows better planning of any negative and possibly disastrous events (storms, droughts and floods). Plausible but invaluable effects, in their turn, may involve timely evacuations of rare species and habitat saving - not to mention the danger of a certain species being washed away by powerful precipitations. This has the potential of assisting in monitoring and mitigating the effects of climate change - abnormal heat or cold spells that are obstructing ecosystems over time in the fluctuating temperature changes. One of the reasons why there are improved weather forecasts is good management of resources. Through the enhancement of irrigation processes, accurate weather and climate projections in the agricultural sector can save them a lot of water, besides the fact that such projections can also contribute towards the saving of this valuable resource in the areas where it is limited.

Such predictions also contribute to reduced use of pesticides and fertilizers which when utilized carelessly cause water and soil pollution. This aim of the research is certainly adjacents to the world efforts to minimize the devaluation of the environment without affecting the productive ability of agriculture.

The accurate weather forecast is essential to the accomplishment of the global environment conservation, including the switching to the renewable energy. The production of solar and wind energy, especially those weather-dependent ones, can be successfully handled in case they are forecasted correctly, thus decreasing the use of fossil fuel power sources and emissions of greenhouse gases. It fosters the mitigation of pollution of the neighboring ecosystems by traditional sources of power and helps in the process of making renewable energy systems more stable. The combination of all these benefits of better weather prediction is very crucial in the survival and maintenance of the ecosystems on the earth as a global human race wisely consumes resources sustainable to them and emits a lesser ecological footprint.

5.3 Ethical Aspects

There is a series of ethical concerns in the use of machine learning methods in weather forecasting; they need to be mitigated in order to apply the techniques responsibly, or to benefit fairly. One of these is data privacy and security as one of the most significant ethical concerns. It is possible that weather prediction systems use vast amounts of resources in form of data generated by a large diversity of sources- personal devices, satellite images, and local sensors. All this information must be gathered, stored and processed in such a way that it does not infringe on individual privacy. People that are developing these systems and users are expected to take into consideration that they are following the data protection policies and they should be honest with people about the way the data is used so that they may be assured that there is no uncertainty in them.

And then, there is another ethical concern; equitable access to proper weather forecasting. Although improved prediction mechanisms become more useful to the larger society, they still may be inaccessible in regions that are strapped financially or socially remote due to lack of money or infrastructure. The advantages may be disproportional and that will result in a digital gap. That way, fairness would need that systems should be availed to all countries despite economical or geographical limitations in a manner that the idea of weather prediction models and its product was availed in a fair way. Such advantages can then be further spread to less privileged groups by liaising with both the government and non-governmental groups to increase it.

Global environmental and social resilience. An ethical obligation on the prevention of weather prediction systems abuse exists as well. The fact that the predictions are highly accurate might be used in a mercenary format, e.g., as a means of speculation in farm produce markets or as a way of providing favoritism to the richer clients as compared to the poorer ones. It needs software creators and interested parties to create models that can achieve ethical usage of these technologies such as open reporting, accountability devices, and rules of regulation.

5.4 Sustanability Plan

We possess a robust sustainability model so that as machine-learning develops in weather prediction, this has to be net-positive in the long run to people and the planet. Thus the sustainability plan under consideration by this research has three principal pillars which are: the ability to keep abreast of new technology, prudent use of resources and joint work with other stakeholders

Technological Adaptability

An adaptation capability that is necessary in supporting relevancy and efficiency of the suggested models of meteorological predictions to continuously-evolving sources of data, variably-configured computing facilities, and solely-evolving climate patterns. The accuracy of the machine learning models will remain constant provided the weather dynamics change provided that they are often retrained using new data. The open-source tools and frameworks are subsequently introduced to the picture in order to enhance the models further by making them even more scalable and friendly to the users so that they can be used in multiple applications around the world. In addition, through the addition of cloud computing and what can be termed edge devices, it would be able to minimize remote reliance on central systems and offer high reliability of operation that is not reliant on location such as at remote locations or unprivileged locations. Multi-Tier Edge-CS Flow of Stream Processing.

Efficiency of Resources

Resource consumption (environmental and computational) Resource efficiency is a large factor of sustainability. Machine learning models, in particular, those, which process a significant amount of data and achieve it with intense speed, are carbon-intensive. It may also be possible to conduct low-environmentaly friendly computational operations by establishing energy-conscious algorithms and providing renewable-energy to green data centers. Such a future models also need to emphasize how it is perceived that data can now be extracted the environment in very green ways, i.e. through sensor nets and public databases--and demonstrate that there is a minimizing aspect to column D in a way that requires not to carry out data collection which is a lot of doing!

Stakeholder Collaboration

Stakeholder Collaboration There are a number of actors involved in sustainability of weather predictions like local communities, companies and governments of the private sector. We can make these models easier to adopt in such ways, and more. useful, and more broadly applicable by working with weather services and environmental groups. We possess an effective long-term model to make sure machine-learning innovation in weather. Adolescence are optimistic to individuals and the environment. Thus, the sustainability plan under development in this research consists of three key pillars, i.e.: the ability to adjust to new technologies, the ability to use the resources judiciously and to cooperate with the stakeholders.

Chapter 6

CONCLUSION AND FUTURE WORK

6.1 Summary of the Study

The present study investigates the use of machine learning methods to improve the accuracy of the forecast for next-day maximum and minimum air temperatures. In this effort, the paper uses the data for observed and forecasted weather over the period 2013-2017 to assess the potential value of selected machine learning algorithms in overcoming the traditional approaches that are inherently problematic when doing weather forecasting. Indeed, the dataset was characterized by varied features, including but not limited to relative humidity, wind speed, cloud cover, and solar radiation, thus forming a strong basis for predictive modeling.

The data pretreatment of the study was essential and approaches such Min-Max Scaling, normalization to scales between[0, 1] or K-Nearest Neighborsfor missing value were performedbefore modeling. Machine learning models such as Linear Regression, Ridge Regression, Decision Tree Regressor,

Random Forest, LightGBM,CatBoost and XGBoost were implemented in the project. R2 Score, Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Error were the metrics used for evaluating the performance of the model. These results showed that ensemble methods such as LightGBM could handlenonlinear relationships and would also emphasize the merits of linear modelsin Linearity of Relationship. LightGBM was the most balancedmodel and it achieved the best performance from these findings.

The paper also examined the broader effects of increased weather forecasting on sustainability, environment and society. It further demonstrated that in case there is good projections, the social benefits in terms ofthe enhanced improvement of agriculture planning and resource allocation, both in disaster preparedness are visible. The resources use was supported with sustainable use, the development of the renewable energy was promoted and the pressure on the environment was decreased significantly.

Outsourcing of these technologies in terms of a sustainability plan and ethical consideration was also provided towards the responsible deployment into the long term. Overall, it was established that machine learning is capable of transforming weather prediction, and therefore providing a solid guiding base to other future researches in the same area.

6.2 Conclusion

The results of the study explain the way the machine learning changes the predictive capabilities of weather forecast systems. Advanced algorithms personally were impressive in the forecast of maximum and minimum air temperatures of the next day, beating by far the traditional forecasting techniques. To allow the prediction models to take advantage of the inherent relation of the features between the timeline and target variable, dataset (the observations made of the physical features of the weather form model forecast) was processed and explored.

Among them, linear frameworks, such as Ridge Regression and Linear Regression were found to be respectable frames used as baselines in air temperature forecasting at a very high accuracy. At the end, though, ensemble techniques, specifically LightGBM, emerged as a compromise of being relatively computationally thrifty, accurate as we can read them, and capturing most of nonlinear interactions. The article stressed the importance of the data characteristics and the needs of the area of implementation determining the approach to adhere to with the ensemble models presented as the more suitable ones with complex data.

The study has in addition to the technical advances, discussed the impact of better weather forecasting to the society, the environment and sustainability. Better predictions result in better control during a disaster, a more efficient allocation of funds, and the ability to include renewable sources of energy to become the ratched contribution of resilience in the climate and to security the environment.

The study also shed some crucial ethical aspects and a sustainability plan that should be put into consideration to guide the responsible usage and implementation of such technologies. The overall conclusion of the research is that machine learning has been crucial in providing solutions to global problems in weather variability and has created more advancements in the research area.

6.3 Implication for Further Study

This has provided an excellent foundation to exploit machine learning in enhancing more precision in weather prediction. However, it gives one a lot of possibilities to explore and be developed on the other side. Protection against such factors as satellite imagery, atmospheric pressure, and sensors as part of the feature set may be considered among the most significant aspects to work on in the future to increase the feature set and enhance the strength of the predictions. This can also permit integration of global datasets to additionally make model generalization that can make predictions in a great variety of geographic areas.

The other potential avenue of development is traditional time-series-based deep learning methods (e.g., LSTM networks and GRUs). In comparison, such models can acquire and encode into them, longer-term dependencies and trends in weather data than classical machine learning algorithms.

The Science Daily article included all the above-said as it was published as a story news article in a recently generated news release on July 20, 2000.

Other follow-up researches can be done on the various sizes and implementation opportunities of weather forecasting systems. This involves, but is not restricted to, creating and adopting energyefficient algorithms to lower the computational price and carbon imaginative impact of large-scale deployments of machine learning. All these models can also be developed in the research to make them user friendly in terms of interfaces and platforms as this allows them to be made available to local communities and farmers as well as policymakers.

With these factors in mind, future work can also lay the premise that any further development of the weather prediction technology will not only be even better technically but also practically and ethically globally.

References

[1] W. Sanders, *Machine Learning TechniquesforWeatherForecasting*, M.S. thesis, Univ. of Georgia, Athens, GA, USA, 2017. [Online]. Available: https://www.ai.uga.edu/sites/default/files/inline-files/theses/sanders_william_s_201712_ms.pdf

[2] M. A. Holmstrom and D. Z. Liu, "Machine learning applied to weather forecasting," 2011. [Online].
Available: https://www.semanticscholar.org/paper/Machine-Learning-Applied-to-Weather-Forecasting-Holmstrom-Liu/e2ed8aba53b4688808d57a0512496beb3548fc2c

[3] P. S. Saketh, R. Rohit, and B. Suneetha, "Weather forecasting using machine learning," in *Proc. Int. Conf. on Intelligent Computing and Communication Technologies (ICICT)*, Apr. 2023, doi: 10.1109/ICICT57646.2023.10134218.

[4] D. Fucci *et al*., "A longitudinal cohort study on the retainment of test-driven development," *arXiv*, 2022.
[Online]. Available: https://arxiv.org/pdf/1807.02971.pdf

[5] M. Kadam, S. Idhate, G. Sonawane, R. Sathe, and P. Gundale, "Weather prediction using machine learning," *Int. J. Creative Research Thoughts*, vol. 11, 2023. [Online]. Available: https://ijcrt.org/papers/IJCRT23A5325.pdf

[6] B. Bochenek and Z. Ustrnul, "Machine learning in weather prediction and climate analyses—Applications and perspectives," *Atmosphere*, vol. 13, no. 2, p. 180, Feb. 2022, doi: 10.3390/atmos13020180.

[7] A. H. M. Jakaria, M. Hossain, and M. A. Rahman, "Smart weather forecasting using machine learning:
A case study in Tennessee," *arXiv*, Aug. 2020, doi: 10.48550/arXiv.2008.10789.

[8] P. Kashyap and R. V. Nayak, "Smart weather prediction techniques using machine learning," 2020.
[Online]. Available:
https://www.irjmets.com/uploadedfiles/paper/volume2/issue_7_july_2020/2600/1628083097.pdf

[9] A. Grover, A. Kapoor, and E. Horvitz, "A deep hybrid model for weather forecasting," in *Proc. 21st ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD)*, 2015, pp. 379–386.

[10] M. Holmstrom, D. Liu, and C. Vo, "Machine learning applied to weather forecasting," 2016.

[11] V. M. Krasnopolsky and M. S. Fox-Rabinovitz, "Complex hybrid models combining deterministic and machine learning components for numerical climate modeling and weather prediction," *Neural Networks*,
vol. 19, no. 2, pp. 122–134, 2006.

[12] F. Montori, L. Bedogni, and L. Bononi, "A collaborative Internet of Things architecture for smart cities and environmental monitoring," *IEEE Internet of Things Journal*, 2017.

[13] Y. Radhika and M. Shashi, “Atmospheric temperature prediction using support vector machines,” *Int. J. Computer Theory and Engineering*, vol. 1, no. 1, pp. 55–58, 2009.

[14] R. K. Yadav and R. Khatri, “A weather forecasting model using the data mining technique,” *Int. J. Computer Applications*, vol. 139, 2016.