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

BY

Toufiqur Rahman Tonmoy
ID: 232-25-011

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

Supervised By

Dr. S. M. Aminul Haque
Professor & Associate Head
Department of CSE
Daffodil International University

Co-Supervised By
Abdus Sattar
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

11TH JANUARY, 2024

APPROVAL

This Project/Thesis titled **"Utilizing machine learning techniques to enhance the accuracy of weather prediction in Bangladesh"**, submitted by **Toufiqur Rahman Tonmoy**, ID No: **232-25-011**, 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 **11-01-2025**.

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 & Director MIS
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

External Examiner

Dr. Mohammed Nasir Uddin, PhD
Professor
Department of Computer Science and Engineering
Jagannath University

DECLARATION

I hereby declare that this research has been done by me under the supervision of **Dr. S. M. Aminul Haque, Professor & Associate Head , Department of CSE,** Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for the award of any degree or diploma.

Supervised by:

Dr. ~~S. M.~~ Aminul Haque
Professor & Associate Head
Department of CSE
Daffodil International University

Co-Supervised by:

Abdus Sattar
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Toufiqur Rahman Tonmoy
ID: 232-25-011
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First I express my heartiest thanks and gratefulness to Almighty Allah for His divine blessing which makes it possible to complete the final year project/internship successfully.

I am really grateful and wish my profound indebtedness to **Dr. S. M. Aminul Haque, Professor & Associate Head,** Department of CSE, Daffodil International University, Dhaka and the deep knowledge & keen interest of my supervisor in the field of machine learning to carry out this project. His endless patience, scholarly guidance, continual encouragement , constant and energetic supervision, constructive criticism , valuable advice, reading many inferior drafts and correcting them at all stages have made it possible to complete this project.

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

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

ABSTRACT

Weather forecasting plays a crucial role in sectors like agriculture, transport, disaster management, and energy. While traditional methods are effective, they often fall short in accuracy due to the complexity of weather patterns. This research explores the use of machine learning to improve next-day maximum and minimum temperature predictions. Data preprocessing, including cleaning and normalization, is vital for maintaining input integrity. Various machine learning models are evaluated against key performance metrics to identify the most effective approaches. Advanced techniques show promise in capturing complex variable relationships, enabling more accurate and actionable forecasts. Improved predictions can reduce disaster risks, optimize agriculture, and support sustainable resource management, including energy and water use. The study also addresses ethical concerns like equitable access, data privacy, and responsible use of technology. This research lays the groundwork for integrating diverse data sources, advanced modeling, and scalable solutions to tackle challenges posed by weather variability. Linear Regression and Ridge Regression did quite well, turning in the highest R^2 Scores of 0.8403 and 0.8402, respectively.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Introduction

Weather prediction is a critical component of modern society, impacting areas such as agriculture, disaster management, aviation, and daily life planning. Fundamentally, these weather forecasts provide the most appropriate risk minimization principle by use of optimized resources for public safety. For example, correct identification of hurricanes or floods or heat waves saves many lives and prevents economic loss.

Numerical Weather Prediction models conventionally form the mainstay for weather forecasting. These models take the help of mathematical equations to simulate the atmospheric behavior based on physical principles. While these models work effectively to some extent, there is a limitation associated with them. They are generally computation-intensive, require huge computing powers, and often fail to reproduce the rapid changes in weather or localized phenomena. Their accuracy is still based on the quality of initial conditions and data inputs alone, and it is subject to errors in various cases.

This becomes possible with recent enhancements in data-driven approaches toward improving weather forecasting. Machine learning provides tools for analysis of historical data on weather, finding hidden patterns in it, and forecasting without explicit knowledge of included physical processes. Such approaches have now been leveraging the growing availability of diverse and large-scale datasets: meteorological observations, sensor data, and satellite imagery [1].

Tapping into these data-driven approaches enhances efficiency and accuracy in weather predictions. This trend towards a shift from strict, traditional physics-based models to data-centered strategies opens new avenues in bettering weather forecasts in ways which have so far not been achievable. The following report assesses how machine learning can overcome traditional challenges to further develop weather prediction.

1.2 Research Motivation

Weather is a vital aspect that helps in shaping human activities and the environment. Pretty accurate weather predictions are essential in many sectors: agriculture, transport, energy, disaster management, and public safety. Farmers depend on weather forecasts to plan planting cycles and harvesting cycles; similarly, transportation systems must depend on weather updates to assure safety and efficiency of the mode, and energy grids need very accurate predictions for handling renewable sources of energy like solar and wind. Reliable and timely forecasts are also crucial in mitigating the effects of very severe weather, such as hurricanes, flooding, and droughts, which have caused massive loss of life and economic damage.

Even with meteorological development, there are still setbacks in the delivery of forecasts that are always reliable. Challenges are amplified in fast-changing weather or critical weather situations, even when small errors can come with grave consequences. Most techniques available today for forecasting poorly capture local weather features or make more accurate short-term predictions. Besides, locations where observational infrastructure is poor develop extra complications in forecast generations due to inadequate information [2].

Because extreme weather events have become increasingly frequent and violent due to climate change, the demand for new and more effective forecasting methods has become even more urgent. Improved weather prediction not only mitigates risks but also forms a basis for pro-active decisions in various enterprises and communities. The opportunity to contribute to this important field by attempting to fill existing gaps in forecast accuracy serves as a strong motivation for the present research.

This work seeks to present new ways of improving weather forecasting, based on increased data availability and advances in data-driven methods. With the aim of overcoming current limitations, this research should lead to more reliable and informative weather forecasts for the good of society and the environment [3].

1.3 Rationale of the Study

Indeed, weather forecasting has always been daunting because atmospheric systems are not only highly complicated but also dynamic. The behavioral pattern of the weather is in turn governed by a host of linked factors such as temperature, humidity, wind patterns, and atmospheric pressure. Therefore, there lies inherently a nonlinear behavior predisposition in the case of weather phenomena and hence is difficult to model. That, however, becomes manifold more complex when it comes to sudden changes like storms, heat waves, and extreme rain showers.

Conventionally, weather forecasting depends chiefly on models of NWP-that is, mathematical simulation of physical atmospheric processes. Such models, while being highly instrumental in the development of meteorology, have some disadvantages. First, their reliance on good-quality initial data necessarily leads to grave errors in forecast results if inaccuracies or gaps are observed within the observational data. Such models do indeed need a great amount of computational resources and are not feasible for real-time applications or local-scale predictions [4].

That is to say, with highly localized or even short-term weather conditions, the conventional means have indeed lost their precision-mostly because there is an acutely felt need now for granularity of data, for capturing sudden variations. Actually, those gaps have appeared in accuracy either for lack of observational infrastructure in some regions or for rapidly changing short-term weather events.

These undertakings would be the keys to better weather forecasts, warnings in good time about extreme weather events, and high-impact benefits to society and economy. New ways of doing the weather forecast, more adaptive, efficient, and accessible, are needed, which can handle intrinsic complexity in atmospheric systems while considering limitations imposed by current practices. It talks of solutions for these kinds of challenges by investigating partial aspects of how the data-driven methodology enhances the accuracy and reliability of the weather forecast.

1.4 Research Question

- Integration of more meteorological data sources would, therefore, improve weather predictions both at various regional and temporal scales.
- What is the critical weakness of the classic method of forecasting the weather, and how might this weakness be overcome in order to enhance the reliability of prediction?
- The inclusion of local weather phenomena would be added in the way of enhancing the preciseness of the short-term forecast of the weather.
- What are the variables that develop errors in the existing forecasting methods, and how could these be minimized in dynamic and rapid weather changes?
- How does the contribution of improved weather prediction systems tie into proactive decision-making in critical sectors like agriculture, transport, and disaster management?

1.5 Expected Output and Objectives

The study falls within the scope of improving accuracy and reliability in weather forecasting-a quite critical area involving many sectors of life such as agriculture, transportation, disaster management, and public safety. Weather prediction is not only dynamic but may be influenced by various factors. Common among them are temperature, atmospheric pressure, and pattern of wind. In that respect, this research has focused on how the limitation of traditional forecasting techniques can be tried through other techniques that go deep into the analysis and interpretation of complicated patterns in weather data.

The aims of the research are:

- **Improvement in the Accuracy of Forecasting**: Devising techniques that would enhance the accuracy of forecast at localized and short-term scales beyond existing capability.
- **Smoothing Gaps in Data**: Better exploitation of available data, taking into consideration real treatments of incomplete or incoherent data sets, which is a common fact in meteorology.
- **Adaptation to Various Scenarios**: Development of forecast models that act optimally for a range of scenarios, from extreme weather events to sparse observation infrastructure over the region.
- **Societal Benefits**: This research will give insight and tools to mitigate these effects of adverse weather events for the preparation and decision-making of individuals and citizens, businesses, and governments.

This research project contributes to narrowing the gap between heightened demands on the accuracy of weather forecasting and deficiencies in what was hitherto available, opening the way for progress which should benefit not only the science as such but society at large.

1.6 Report Layout

In Chapter 1, the introduction, objectives, and key research inquiries of the study are outlined. In Chapter 2, concise synopses of the literature review are provided. In Chapter 3, the proposed methodology is described in detail. In Chapter 4, the experimental outcomes of the paper are described and examined. The fifth chapter discusses the sustainability plan, societal and environmental impacts, and ethical considerations. The sixth chapter concludes the present investigation and outlines a strategy for subsequent endeavors.

CHAPTER 2

BACKGROUND

2.1 Preliminaries/Terminologies

Within the last few decades, weather forecasting has been one of the most studied topics since it impacts almost all aspects of human life and nature. To mitigate the dire effects of harsh weather conditions, maximize efficiency in resource use, and ensure the safety of the public, more accurate weather forecasts are necessary. Throughout many years, improvements in weather forecasting have been proposed using various techniques, starting from the traditional numerical models up to the recent advanced data-driven approaches [5].

Literature on weather forecasting has indicated the complicated and dynamic nature of the atmospheric system, which always creates difficulties. While traditional methods of forecasting, based on mathematical simulations of atmospheric physics, have reached significant milestones, nonlinear interactions and sharp changes in weather patterns remain a limitation to these methods. Localized and real-time demands for meteorological forecasts expose conventional approaches as deficient in computational efficiency and precision.

New paths toward better weather prediction have become possible in recent years due to the increase in available data and processing capability. Various research has explored various sources of data, including satellite imagery, sensor networks, and historical records of weather, in order to obtain stronger and adaptive forecasting methods. 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.

2.2 Related Works

Holmstrom et al. gave a predictive model for the maximum and minimum temperatures for the forthcoming seven days based on input from the last couple of days [10]. Using linear regression, Holmstrom et al. gave a predictive model and extension through the so-called functional linear regression model. While these models showed some success, they fell far short of the accuracy obtained by professional weather forecasting services for the seven-day forecast. The models were far superior in longer-range forecasts or even in determining trends. In the weather beyond the seven-day forecast period.

Another work by Krasnopolsky and Rabinovitz made an attempt in another direction. Their model is also hybrid, where the neural networks again take center stage in an effort to model the physics behind the weather phenomena [11]. Their approach focused on using the computation powers of neural networks to increase the accuracy of the forecast of weather by emulating the complex dynamics of the atmospheric state.

On the other hand, Radhika et al. treated weather forecasting as a classification problem and used support vector machines to predict the weather conditions based on the pattern observed within the historical data itself [13]. This approach indeed explored the usability of machine learning in weather forecasting and was quite effective in categorizing and providing identifications concerning specific states.

In [14], there is a proposal of a data mining-based prediction framework that can detect fluctuating trends in historical weather data. In the approach, a Hidden Markov Model is considered; it further integrates k-means clustering in order to extract meaningful observations from historical data regarding 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. investigated a hybrid model that uses discriminatively trained predictive models in conjunction with deep neural networks to predict weather forecasts [9]. The model specifically targeted the joint statistical dependencies across a panel of weather indicators to allow for nuanced modeling of how these indicators change over time together. In the integration of these predictive tools, their model was more accurate at forecasting, particularly complex weather scenarios.

Another innovative approach was proposed by Montori et al. The authors developed a crowdsensing-based approach, meant to provide environmental phenomena monitoring through voluntary sharing of data collected by individual users' smartphones [12]. In this context, they designed an architecture called SenSquare, aimed at aggregating data from IoT sources and crowdsensing platforms. The collected data was represented in a harmonized format, accessible for applications like smart city environmental monitoring. The use of crowdsensed data introduced a new paradigm in monitoring and forecasting weather due to the nature of real-time environmental data shared by participants.

Different as these approaches were-from those that employed machine learning models to hybrid systems and further to crowdsourced data-no study actually indicated the integration of data from the neighboring geographical areas to improve the accuracy of weather prediction. The lacuna in this respect provides scope for future studies to investigate how spatial data from contiguous regions could be used to augment the performance of the forecasting.

2.3 Problem Scope

Weather forecasting is an inherently complex task due to the highly dynamic and nonlinear nature of meteorological phenomena. Relations which are not easily captured through more traditional statistical techniques across longer horizons when applied to historical data analysis. While professional forecasting services are 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 consider spatial 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 approaches where physics-based models are coupled with machine learning techniques have been able to achieve certain potentiality for bridging this gap. For example, neural networks are good at modeling nonlinear dynamics developed by weather systems. Still, most of these models rely so much on historical data without considering real-time information from surrounding areas. The fragmentation of such data in this manner seriously limits their adaptiveness in case of sudden changes in weather-such as storms or changes in temperature-which are really crucial for effective forecasting.

While techniques such as data mining and probabilistic models, for example, Hidden Markov Models, have been applied in the identification of repetitive patterns within historical weather data, most techniques developed typically lack in dealing with real-time dynamic situations. Similarly, other clustering methods, such as k-means, are inherently slanted toward the detection of trends rather than the extraction of temporal and spatial complexities obvious in forecasts of the weather. These limitations thus create a case for developing models that can merge insights from historical data with real-time and spatially informed data feeds [7].

These are challenges that might be overcome with the help of emerging technologies in crowdsensing and IoT-based environmental monitoring. Such systems are able to gather real-time information from a network of devices and users, providing a continuous stream of rich, localized weather information. The current practical applications using such technologies have a greater focus on environmental rather than direct weather monitoring. Merging this with predictive modeling, especially when combined with data from other surrounding areas, may give a path to even more advanced and improved forecasting systems.

2.4 Challenge

The task of weather prediction has a number of inherent difficulties because the atmospheric system represents a dynamic system with a very complex interaction between its different elements. One of the major challenges is the nonlinear relationships that exist among various meteorological variables, which include temperature, humidity, wind speed, and atmospheric pressure-all these variables are interdependent with external factors, such as solar radiation and topographical features.

There is, again, the issue of data availability and quality. Weather forecasting reliably requires large amounts of high-resolution data gathered over a certain period of time from various sources. Unfortunately, inconsistencies, gaps, and noise in observational data can reduce model prediction accuracy. Real-time data gathering can still be quite a challenge from remote or hard-to-reach areas, and that puts limits on the models' ability to predict with greater resolution at those locations.

Another important challenge is incorporating spatial dependencies into weather forecasting models. Atmospheric phenomena are usually connected across space, where conditions in one area may influence the patterns in neighboring areas. The general focus of most existing models is on isolated data that do not effectively utilize this spatial information, which leads to not-so-accurate regional forecasting. It requires advanced techniques and much more computational resources to integrate data from neighboring locations, which becomes challenging when dealing on a large scale [8].

Model scalability and computational complexity are also paramount challenges. Most current machine learning models, particularly deep learning architectures, require a great deal of computational power and memory during training and deployment. For instance, hybrid models that integrate physical-based simulation with machine learning algorithms require intensive computation; therefore, real-time forecasting can hardly be achieved. How to balance model accuracy with computational efficiency remains a stubborn challenge.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Proposed Methodology

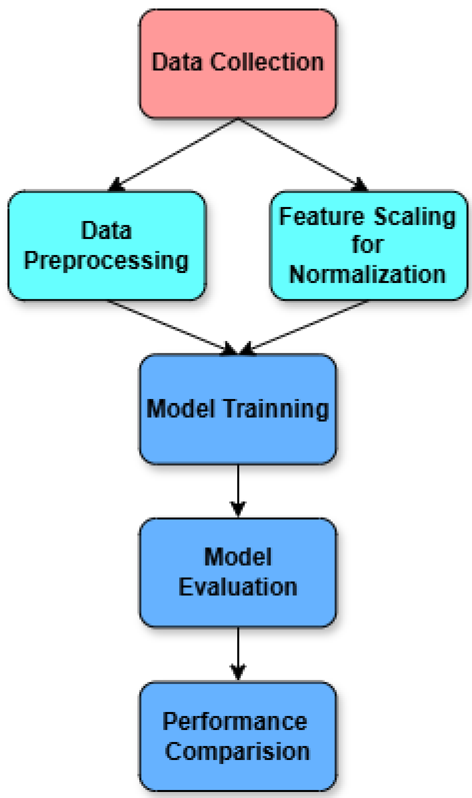

Figure 3.1: Proposed Research Methodology

The methodology to be followed in this research will emphasize the elaboration of a robust framework in predicting the next day's maximum and minimum air temperatures using machine learning techniques. In data preprocessing, extensive handling of missing values through imputation will be done with the K-Nearest Neighbors (KNN) Imputer to ensure consistency in the data with preservation of variable relationships. This dataset contains observation and forecast information about the weatherfrom 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 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 predictions of 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 methods such as XGBoost, LightGBM, and CatBoost. Some of the performance metrics used in the analysis include the R^2 score, mean squared error, root mean squared error, and mean absolute error. The methodology utilized in this approach was such that through multi-output regression, it was able to identify interdependence between target variables with the express aim 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.

Present_Tmax vs Next Tmax:

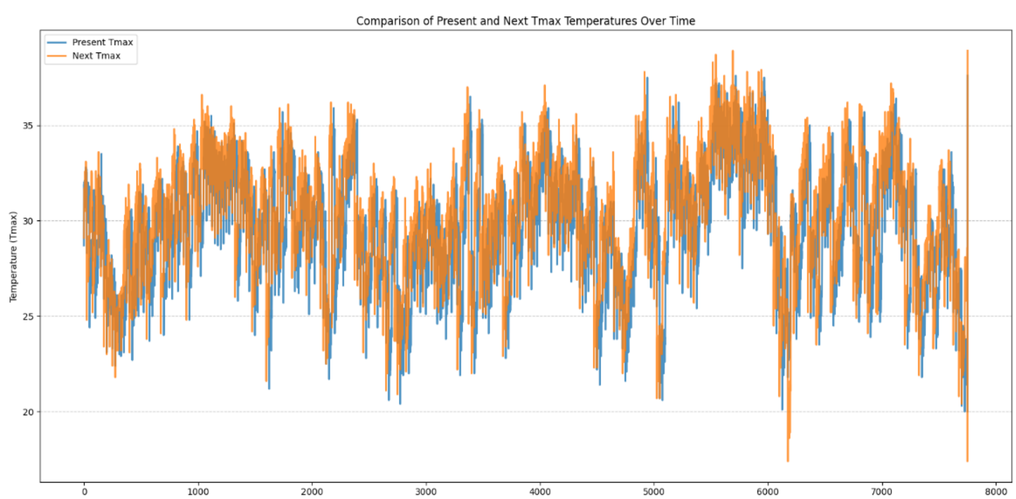

Figure 3.2: Present_Tmax vs Next Tmax

Tmin vs Next Tmin:

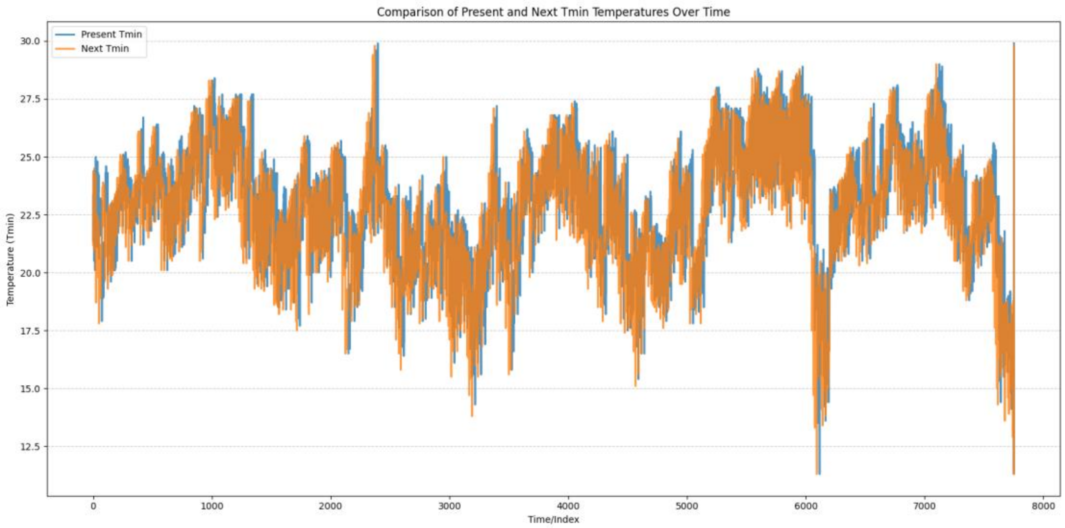

Figure 3.3: Tmin vs Next Tmin

3.2 Data Cleaning

Data cleaning is an indispensable part of any data-driven project, which involves the handling of real-world datasets, such as weather prediction data. Some of the features in this project's dataset contained missing values that could seriously affect the performance and reliability of machine learning models. Here is how the missing values were dealt with in this study:

Missing Value Identification

- First of all, the dataset was loaded and then it intensely explored to missing value imputation was executed for the dataset. It has been executed using .isna().sum() will give the count of null values across each column.
- Some columns like station and Date have been dropped as they do not contribute to the predictive modeling And identifiers instead of features.

Imputation Technique

1. K-Nearest Neighbour imputer is considered one of the potent ways of handling missing values through imputation based on the similarity of data points. The K-Nearest Neighbour imputer works by using the concept of a neighborhood of data points in order to predict values of missing variables, considering only those imputed values which would go in accordance with the distribution of the already existing data.

2. The KNN imputer is initialized with 2 neighbors as a trade-off between accuracy and computational efficiency.

3. Imputation has been done for all the numerical columns without the target variables, Next_Tmax and Next_Tmin, to maintain the integrity of the predictive modeling.

Benefits of Using KNN Imputation

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

3.3 Feature Scaling for Normalization

Feature scaling is a critical preprocessing step to ensure that all features contribute equally to the predictive modeling process. Weather prediction data typically involves features with vastly different ranges, such as temperatures (in °C), wind speed (in m/s), and solar radiation (in Wh/m^2). Without scaling, models might give undue weight to features with larger magnitudes, leading to biased predictions.

Why Min-Max Scaling?

- **Min-Max Scaling** was chosen as the scaling technique for this project. It transforms the data into a specified range, typically [0, 1], making it ideal for algorithms that rely on distance metrics (e.g., regression, decision trees).
- The formula for Min-Max Scaling is:

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

Where:

- X: Original feature value
- Xmin: Minimum value of the feature
- Xmax: Maximum value of the feature
- Xscaled: Scaled value within the range [0, 1].

Steps in Feature Scaling

- The dataset was processed to remove non-numeric columns, such as station and Date, as these were not numerical and did not require scaling.
- The MinMaxScaler from the sklearn.preprocessing library was initialized and applied to all features in the dataset.
- Both input features and output variables (Next_Tmax and Next_Tmin) were scaled separately to ensure compatibility with the machine learning models.

Advantages of Scaling

- Ensures all features are on the same scale, preventing dominance by features with larger ranges.
- Speeds up the convergence of gradient-based optimization algorithms.

- Enhances the performance of machine learning models, particularly those sensitive to feature magnitudes (e.g., decision trees, ensemble methods, and neural networks).

3.4 Multi-Output Regression for Simultaneous Prediction

Predicting both Next_Tmax and Next_Tmin simultaneously required a multi-output regression approach. Multi-output regression is a supervised learning technique where multiple dependent variables are predicted from a single set of input features. This approach was suitable for weather prediction as Next_Tmax and Next_Tmin are closely related and share common predictive features.

Why Multi-Output Regression?

- **Correlated Outputs:** Maximum and minimum temperatures often have a natural correlation due to shared influencing factors, such as humidity, solar radiation, and wind speed.
- **Efficiency:** Instead of building separate models for each target variable, a single multi-output regression model can handle both predictions simultaneously, reducing computational overhead.

The MultiOutputRegressor class from the sklearn.multioutput library was used to extend single-output regressors into multi-output regressors. Various base regressors were tested as the underlying models:

- **Linear Regression:** To establish a baseline performance.
- **Ridge Regression:** To incorporate regularization and prevent overfitting.
- **Decision Tree Regressor:** To capture non-linear relationships.
- **Random Forest Regressor:** To improve accuracy through ensemble learning.
- **XGBoost:** To leverage gradient boosting for enhanced performance.

- **LightGBM:** For its computational efficiency and accuracy.
- **CatBoost:** To handle categorical features and improve performance.

Model Training and Evaluation

1. **Training:** Each model was trained on the scaled training dataset using the MultiOutputRegressor. The inputs were the scaled features, and the outputs were Next_Tmax and Next_Tmin.
2. **Prediction:** Predictions for the test dataset were generated for both target variables.
3. **Evaluation Metrics:**

   R^2 Score: Measures the proportion of variance explained by the model.

   Mean Squared Error (MSE): Quantifies the average squared difference between predicted and actual values.

   Root Mean Squared Error (RMSE): Provides an interpretable metric in the same unit as the target variable.

   Mean Absolute Error (MAE): Measures the average magnitude of errors.

Multi-output regression exploits the interdependence of the target variables to improve prediction accuracy. Reduces redundancy by training a single model instead of two independent ones.

3.5 Algorithm Description

Linear Regression

Linear regression is a foundational algorithm in supervised learning and was used in this study as a baseline model. It works by modeling the relationship between independent variables (features) and dependent variables (targets) using a linear equation:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon$$ -----------(2)

Where Y is the target variable, $X_1, X_2, \ldots, X_n$ are the predictors, $\beta_0, \beta_1, \ldots, \beta_n$ βare the coefficients, and ϵ represents the error term. Linear regression assumes a linear relationship between predictors and targets, making it interpretable and computationally efficient. However, its simplicity can be a limitation, especially in capturing complex non-linear relationships. Despite this, it serves as an essential starting point to gauge the baseline performance of the dataset.

Ridge Regression

Ridge regression builds upon linear regression by adding a regularization term to the cost function to prevent overfitting:

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

Where λ is the regularization parameter, and β_j are the coefficients. By penalizing large coefficients, Ridge regression reduces model complexity and enhances generalization to unseen data. It is particularly useful for datasets with multicollinearity, as it reduces the variance of the estimates. This makes Ridge regression a robust alternative to linear regression, particularly for handling noisy or highly correlated data features in weather prediction.

Decision Tree Regressor

The threshold at the nodes splits the data within the tree structure to predict continuous target variables by minimizing a loss function, usually the mean squared error. Decision trees capture nonlinear relationships and interactions between features very well. Interpretability is considered its major strength since the rules of predictions are easily visual and understandable. However, decision trees are susceptible to overfitting, especially with deep trees, which was mitigated by the ensemble methods themselves in this research.

Random Forest Regressor: Random Forest is a type of ensemble learning where generalization is done by aggregating multiple decision trees. This actually reduces overfitting. In this process, each tree is trained on a random subset of the training data. Predictions are combined; most of the time, it averages on regression tasks. Due to the averaging effect inside the ensemble, the Random Forest is generally resistant to overfitting. This is a model that can handle high-dimensional big data with ease. The result also gives feature importance scores that are useful to understand the contribution of each feature toward the model's predictions. It works out pretty effectively on nonlinear data distributions, so that will be a good fit for weather prediction.

XGBoost (Extreme Gradient Boosting):XGBoost is indeed impressively effective in gradient boosting, building ensembles of trees sequentially, trying to correct at every step the errors of the model built so far. Success here is based on optimization, thanks to a customized objective function employing second-order gradients, therefore guaranteeing better results in terms of both accuracy and efficiency. It will also include the regularization parameters, which will avoid overfitting, and support the distributed computing that will enable faster training on big data. Generally, XGBoost has wide adaptation due to its flexibility and scalability, suitable for complicated datasets, especially those concerning weather prediction, where the relationships between features are intricate and non-linear.

LightGBM (Light Gradient Boosting Machine)

LightGBM is an efficient and fast gradient boosting algorithm. Unlike general Boosting Algorithms, it makes use of a histogram-based approach for finding the best splits, which drastically reduces the consumption of memory and training time. It also has a set of techniques such as leaf-wise tree growth that reduce errors more efficiently than conventional level-wise growth, hence being quite efficient in big-scale datasets and high-dimensional data, say like in weather prediction. This goes further in extending real-world applicability into its capability of handling sparse data and categorical features.

CatBoost

CatBoost is intrinsically a gradient boost algorithm with a twist for natively handling categorical features directly without any type of extensive preprocessing, such as one-hot encoding. This might employ some interesting technique called ordered boosting ensembled with oblivious trees, combined to present state-of-the-art results with least overfitting. On average, CatBoost proves to be very efficient and mostly outperforms other algorithms which are based on gradient boosting for datasets containing both numerical and categorical features. It worked particularly well in this research because it captured even very complex interactions among the weather variables with no great loss of computational efficiency.

CHAPTER 4
EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Overview of Model Evaluation Metrics

The following section summarizes all the applied evaluation metrics to determine the performance. These include the R^2 Score, Mean Squared Error-MSE, Root, Mean, Squared Error-RMSE, and Mean Absolute Error-MAE, within the models set against the next day's maximum Next_Tmax and minimum temperatures Next_Tmin. The R^2 Score describes the percentage of variance in a dependent or target variable that has been captured or described by the model. A higher value toward 1 indicates better performance. MSE, RMSE, and MAE all provide error measures in different scales and hence enable both statistical and practical insight from the models' predictions. With that said, the stage is now set to delve into the details of individual model performances and various strengths and weaknesses comparative to one another.

4.2 Experimental Results & Analysis

	Model	R2 Score	Mean Squared Error (MSE)	Root Mean Squared Error (RMSE)	Mean Absolute Error (MAE)
0	Linear Regression	0.840266	1.843438	1.357733	0.996920
1	Ridge	0.840263	1.843230	1.357656	0.996871
5	LightGBM	0.796422	1.921213	1.386078	0.998687
3	RandomForest Regressor	0.770832	2.102042	1.449842	1.065615
6	CatBoost	0.770075	2.013041	1.418817	1.018908
4	XGBoost	0.743892	2.265336	1.505103	1.090621
2	DecisionTree Regressor	0.691186	3.459738	1.860037	1.399312

Table 4.1: Model Evaluation Table

The performance comparisons of the two machine learning models done for high and low temperatures of the next day would consider that each algorithm would have some strengths and weaknesses. In this case, performance comparison and evaluation against each other are considered by metrics such as R^2 Score, Mean Squared Error-MSE, Root Mean Squared Error-RMSE, and Mean Absolute Error-MAE.

Of these, Linear Regression and Ridge Regression had the best overall results: an R^2 Score of 0.8403 and 0.8402, respectively. Both models were able to account for approximately 84% of the variance within their target variables and showed strengths in their modeling of linear associations among features and outputs. Their MSE, RMSE, and MAE values were the lowest compared to the rest of the models, which means they should have been very consistent to make a wide range of predictions.The negligible differences between these two models underpin the limited impact of regularization in Ridge Regression for this dataset.

LightGBM emerged as the most effective ensemble method, achieving an R^2 Score of 0.7964. While slightly less accurate than linear models, it excelled in capturing non-linear patterns in the data, evident in its competitive RMSE of 1.386 and MAE of 0.999. Its histogram-based approach and efficient leaf-wise tree growth allowed it to outperform other ensemble methods like Random Forest and CatBoost. While LightGBM remains strong in its non-linear modeling, the error metrics are slightly higher compared with the linear models to highlight the dataset challenges in capturing extreme values.

Random Forest and CatBoost performed similarly, by posting 0.7709 and 0.7701 R^2 Scores, respectively. Random Forest was better by its ensemble averaging approach and reduced the overfitting and variance, thus allowing it to reach an RMSE of 1.456 and MAE of 1.07. While CatBoost was able to make use of the gradient boosting technique and ran with an RMSE of 1.419 and MAE of 1.019, making it slightly better than Random Forest. Both models showed promise in this competition but struggled to match the precision of LightGBM in terms of overall accuracy.

XGBoost reached an R^2 Score of 0.7439, ranking lower in the ensembles. Its RMSE was 1.505, and MAE was 1.091, which reflected its limitations in handling this dataset compared to LightGBM and CatBoost. While XGBoost remains an extremely popular choice for many tasks of prediction, this finding from the study indeed shows that its default configurations may require further optimization to excel in weather prediction tasks.

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 emphasize its 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.

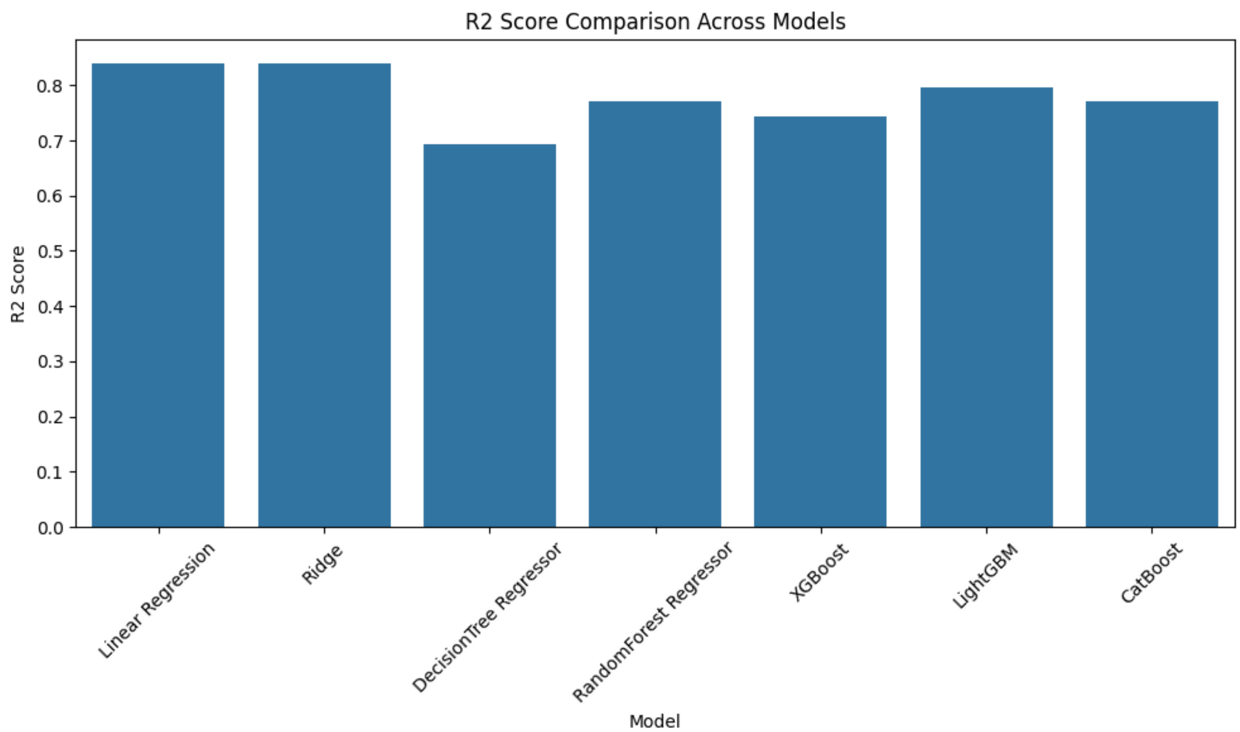

Figure 4.1: R2 Score Comparison Across Models

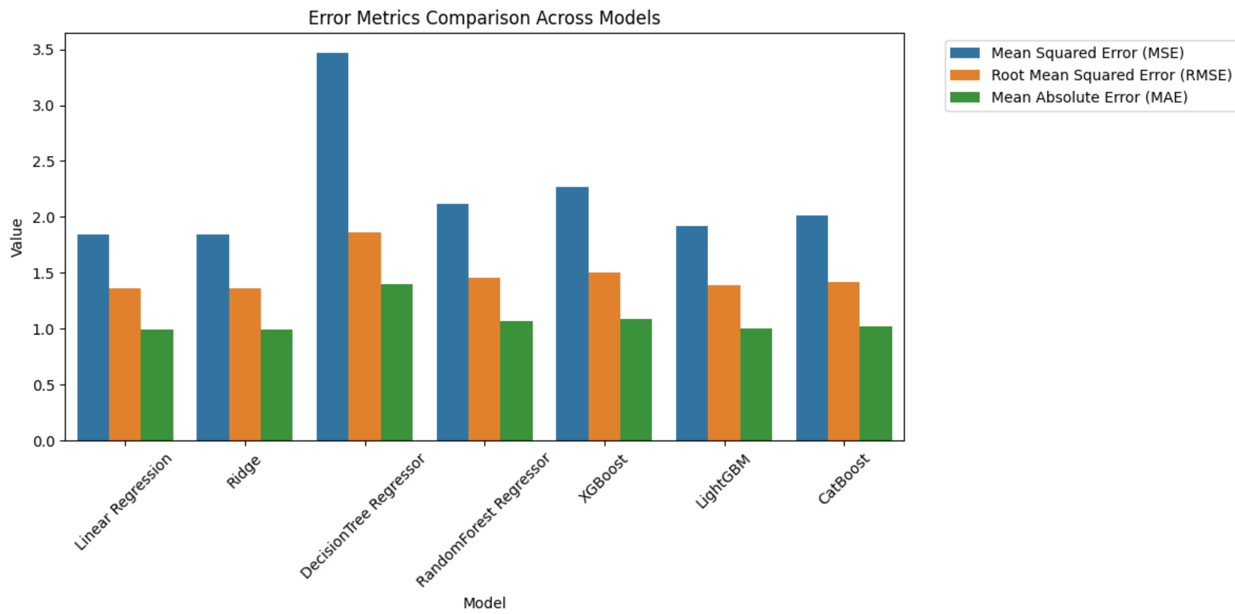

Figure 4.2: Error Metrics Comparison Across Models

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

Ensemble learning models such as Random Forest Regressor, LightGBM, XGBoost, and CatBoost were investigated for their ability to pick up complex nonlinear relationships in the dataset. Considering these, LightGBM showed the highest with an R^2 Score of 0.796, trailed by Random Forest at 0.771, followed by CatBoost at 0.770 and XGBoost with a score of 0.744. Its histogram-based approach, along with its leaf-wise growth strategy, may explain LightGBM's efficiency and accuracy. Random Forest fared well, benefiting from averaging out several of the decision trees which reduced overfitting and variance. CatBoost had reasonable performance, handling categorical features nicely and falling short of LightGBM. Similarly, XGBoost and its gradient-boosting techniques fell short. These results show that these ensemble methods bring significant improvement in predictive accuracy compared to linear models but at a certain computational cost.

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

Understanding the errors made by each model is critical for identifying areas of improvement. While Linear and Ridge Regression generally recorded overall low general error metrics, they strongly struggled with the nonlinear pattern, especially for the extreme values of Next_Tmax and Next_Tmin. Ensembling models do considerably better at these extremes, while LightGBM and CatBoost sometimes generate larger errors on the moderate values that could indicate overfitting to some patterns in the data. This section also gives the discussion on implications for the values of MAE ranging from 0.996 for Ridge up to 1.39 for Decision Tree Regressor, therefore highlighting the practical importance of these errors in real-world weather prediction.

4.7 Discussion

This analysis of model performance gives a comprehensive overview of the different machine learning algorithm predictions on the maximum and minimum air temperature for the next day. Linear Regression and Ridge Regression did quite well, turning in the highest R^2 Scores of 0.8403 and 0.8402, respectively. In fact, they are very strong at capturing, through linear relationships, the variance in the datasets. The error metrics, including RMSE and MAE, are low, relating to their reliabilities as baseline models. However, simplicity confines them to nonlinear dynamics in data, hence applicability in complicated patterns capturing is low regarding weather. Because of the regularization capability of Ridge Regression,

This represents very minor improvements over Linear Regression and thus suggests that the dataset suffers neither from significant multicollinearity nor overfitting issues.

Among ensemble models, LightGBM proved to be the most efficient with an R^2 Score of 0.7964, proving its powers in nonlinear relationship modeling with high-dimensional data. This stands tall and competent concerning accuracy and efficiency for practical deployment when compared with other ensemble methods such as Random Forest and CatBoost. While Random Forest and CatBoost showed very close results, featuring respective R^2 Scores of 0.7709 and 0.7701, their error metrics were higher than those of the other two models, meaning that further tuning might be necessary for the two algorithms. XGBoost, usually robust, underperformed in this study, reaching an R^2 Score of 0.7439, probably due to unsuitable default settings. Finally, the Decision Tree Regressor, with the lowestR^2 Score of 0.6922, struggled to generalize effectively, reflecting its limitations as a standalone model for complex prediction tasks. These findings highlight the trade-offs between interpretability, computational efficiency, and predictive accuracy, with LightGBM emerging as the most balanced and robust option for weather prediction in this study.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 Impact on Society

The integration of machine learning techniques into meteorological forecasting could have some radical positive impacts on society. Accurate weather forecasting plays a critical role in safeguarding lives, property, and livelihoods. By enhancing the precision of predictions for extreme weather events, such as heatwaves, cold spells, or sudden temperature fluctuations, this research will enable communities to better prepare for and mitigate the risks associated with climate variability. Early and accurate forecasting by the meteorological department enables different government agencies and disaster management groups to initiate necessary precautions well in advance through warnings, evacuation plans, or mobilization of resources to mitigate adverse effects of weather hazards.

Furthermore, better weather forecasting has huge applications in highly weather-sensitive sectors like agriculture, transportation, and energy. Precise temperature forecasts allow farmers to determine the best planting schedules, irrigation, and pest control methods, thus maximizing crop yield with minimal loss. On the other hand, transport industries like aviation and shipping could improve safety and efficiency in operations by making use of accurate temperature and weather forecasts. In the energy sector, renewable energy providers can use forecasts in managing solar and wind energy generation efficiently to reduce reliance on non-renewable sources of energy. From this perspective, the impact of this research on society goes beyond questions of immediate safety to include economic resilience, optimization of resources, and better quality of life.

5.2 Impact on Environment

Application of machine learning to improve the accuracy of weather forecasting will surely have huge positive impacts on the environment by allowing proactive environmental management. Accurate and precise weather forecasting helps prepare for adverse events like storms, droughts, and floods, which generally result in widespread environmental degradation. For example, accurate forecasts enable the relocation of endangered species, Their habitat protection, solution of soil erosion problems when heavy rain takes place, would be very timely. Long-term temperature changes can be of great use in order to monitor and mitigate those disturbances in climate change-like unusual heat waves or a long-lingering cold spell-that disturb an ecosystem in a very upsetting way.

Good weather forecasting largely concerns proper resource management. With a proper forecast of say temperature and rainfall for agricultural activities, there would be no wastes of water through irrigation done at appropriate times, hence conserving this important resource in places where it is particularly short. Better forecast reduces unnecessary application of fertilizers and pesticides which often when applied in excess harm the soil and water systems. In this view, the current research fits into the global effort at ensuring minimal environmental degradation without necessarily affecting agricultural productivity. Precise forecasting of renewable energy-especially considered a milestone in most environmental sustainability matters around the world-is increasingly bolstering the switch to renewables. Solar and wind energy productions are highly connected with favorable climatic conditions; hence, precise forecasting assures effective management of the energy grid with less dependence on fossil fuels and reduced levels of greenhouse gas emissions. This will help in improving the reliability of renewable energy systems and decreasing environmental pollution due to conventional energy sources. Improved forecasting would help protect the ecosystems, furthers the good use of natural resources, and reduces the ecological footprint of mankind

5.3 Ethical Aspects

A few ethical challenges arise with the use of machine learning techniques in weather forecasting and need consideration with the aim of achieving its responsible use and or equitable benefit. Among major ethical considerations of this regard, privacy and security of data are an issue. For example, weather prediction systems can base their operation on vast resources emanating from personal devices, satellite imagery, and local sensors. It is in this regard that such information should be collected, stored, and processed in a manner which would not infringe on personal privacy, and any data protection policies regarding such detail would be duly implemented and transparency displayed regarding the usage of such information to gain the confidence of the public in general.

That creates yet another ethical issue: fair access to good weather forecasting. Improved forecasting systems serve the interests of the public, but this may be prohibited by infrastructural or fiscal constraints in economically or geographically challenged regions. This could lead to some form of a digital divide where benefits may fall very unevenly. By fairness in this sense, equal access to the models of weather prediction and their outputs nondiscriminatorily among the geographically or economically constrained countries will be provided. Collaboration for these benefits with not only governmental but also nongovernmental organizations provides an avenue toward the accomplishment of societal and environmental resilience at large.

There is also an ethical obligation with respect to the risk of abuse of the weather forecasting systems. It is possible that such climate forecasts could be used for mercenary ends, such as speculation in potato commodities, heavy land speculation, or preferential treatment of the affluent over the poor. It necessitates the developers and other stakeholders to come up with frameworks that can help minimize the abuse of the technologies including how they are reported, how they are implemented, who is held accountable for them and the policies controlling their application. Shifting focus to these ethical dilemmas means this research has to implement the deployment of machine learning in weather prediction that seeks to assist society but upholds the requirements of fairness and equity as well as responsibility.

5.4 Sustainability Plan

An effective sustainability strategy is being implemented to guarantee that advancements in artificial intelligence on weather forecasting will be of benefit to society and the environment in the future. The sustainability plan of this research is founded on three pillars: technological adaptability, resource efficiency, and stakeholder collaboration.

Technological Adaptability

In order to enhance the life span and effectiveness of the recommended models of weather prediction, it is imperative that there is responsiveness to new sources of data, changes in capabilities, and trends in climate within the global context. Machine learning performs well on models that are trained with recent data. Regular data updates may even assist in improving an exported model's performance during changing weather conditions. The provision of additional resources such as tools and frameworks that are open-source is also expected to enhance the models' usability on different applications worldwide scale and increase their scope. Cloud computing and edge devices combined can also minimize the need for centralized systems, which enables dependable operations in rural or underdeveloped areas to be achieved.

Efficiency of Resources

The efficiency of the cyber environment is an important factor for environmental sustainability. Machine learning models, especially those involving big data and heavy computing, can have a large carbon footprint. The environmental impact of computing systems can be minimized as much as possible by optimizing algorithms for energy consumption and green data center optimization for renewable energy users on the snow Besides, the models should focus on sustainable data acquisition methods, for instance, making use of publicly available data sets and sensor networks, to decrease the demand for resource-intensive data acquisition procedures as much as possible.

Stakeholder Collaboration

Sustainability in the realm of weather prediction involves a number of agents: governments, private organizations, and local communities. Partnerships with meteorological agencies and environmental organizations can facilitate the integration of the models into existing systems, enhancing utility and reach. This would involve dissemination and training on the use of forecasts at the community level to ensure the benefit trickles down to the grassroots level. Mechanisms of funding, through public-private partnership or international grants, could also be developed to ensure the continuation and scaling in resource-constrained settings.

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.

Data preprocessing in the study played an important role; thus, techniques such as K-Nearest Neighbors imputation for handling missing values and Min-Max Scaling for normalization were performed in preparation for modeling. Various machine learning algorithms were applied in the experiment, including Linear Regression, Ridge Regression, Decision Tree Regressor, Random Forest, LightGBM, CatBoost, and XGBoost. The model performance was measured using evaluation metrics: R^2 Score, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error. These results pointed out the strengths of linear models in the Linearity of Relationships and how ensemble methods, such as LightGBM, could handle nonlinear patterns; LightGBM was the most balanced and effective model in the results.

The study also looked at the broader impacts of better weather forecasting in terms of society, the environment and sustainability. It showed that accurate forecasting has a positive impact on social benefits, such as disaster preparedness, improved agricultural management and resource efficiency. The environmental impact was profound, renewable energy management supported, and sustainable resource use ensured. Ethical consideration and a sustainability plan were also given out for the responsible deployment and feasibility of these technologies in the longer run. All in all, it was proved that machine learning can revolutionize weather forecasting, hence giving a good foundation for other future studies in this field.

6.2 Conclusions

The findings of this study highlight the transforming impact of machine learning on improving the accuracy and reliability of weather prediction systems. Advanced algorithms applied to predict next-day maximum and minimum air temperatures significantly outperformed traditional forecasting methods. The dataset, consisting of weather observations and model forecasts, was preprocessed and analyzed efficiently such that the predictive models could make use of the inherent relationships between the features and target variables.

Linear models were among the best, including Linear Regression and Ridge Regression, which constituted very robust baselines for high-accuracy air temperature predictions. However, ensemble methods, particularly LightGBM, proved to be the most effective in balancing predictive accuracy, computational efficiency, and capturing most nonlinear relationships. The investigation emphasized that the choice of algorithm should be guided by the nature of the data and the specific requirements of the application, where ensemble models were more suitable for complex datasets.

Besides technical developments, the study discussed the implications of improved weather forecasting for society, the environment, and sustainabilityAdvanced forecasting allows better risk management, improved productivity, and the integration of renewable energy to further contribute to resilience and environmental protection in society Research also highlights many important ethical considerations and sustainability policies that should be considered to determine the responsible production and maintenance of these emphasizes technical skills. The final conclusion of the study was that the device learning has played a vital role in solving global challenges in weather variability and has opened more innovations in the field of study.

6.3 Implication for Further Study

This provided a very good foundation for applying machine learning in improving the accuracy of weather forecasting. In other words, it opens up many possibilities for exploration and improvement. In future research, adding features such as satellite imagery, atmospheric pressure data, and sensors to the inventory could enhance it and increase the robustness of forecasts This, in turn, could provide global data a list has been added to greatly facilitate model generalization that can predict across geographic areas.

Another promising direction is the use of deep learning techniques, such as LSTM networks and GRUs, designed for time-series data. These models are better able to identify and capture the long-term dependence and dynamics of climate data compared to classical machine learning algorithms Additional guidelines in addition to deep learning algorithms may lead to exploration depth in mixed model results from complex ensemble methods You can focus on the unique scale and weather forecasting systems that can be used.This includes, but is not limited to, the development and implementation of energy-efficient algorithms to reduce the computational costs and carbon footprint from large-scale applications of machine learning.

The research may also focus on developing user-friendly interfaces and platforms so these models can become accessible to local communities, farmers, and policymakers. By considering these aspects, future work can also ensure that any furtherance of weather prediction technology is not only superior technically but also practically and ethically on a global scale.

REFERENCES

[1] W. Sanders, "MACHINE LEARNING TECHNIQUES FOR WEATHER FORECASTING." Available: https://www.ai.uga.edu/sites/default/files/inline-files/theses/sanders_william_s_201712_ms.pdf

[2] M. A. Holmstrom and Dylan Zhe Liu, "Machine Learning Applied to Weather Forecasting," 2011. https://www.semanticscholar.org/paper/Machine-Learning-Applied-to-Weather-Forecasting-Holmstrom-Liu/e2ed8aba53b4688808d57a0512496beb3548fc2c (accessed Jan. 04, 2020).

[3] Podicheti Sriram Saketh, Relangi Rohit, and B. Suneetha, "Weather Forecasting using Machine Learning," Apr. 2023, doi: https://doi.org/10.1109/icict57646.2023.10134218.

[4] D. Fucci *et al.*, "A Longitudinal Cohort Study on the Retainment of Test-Driven Development," *arXiv*, vol. 1, no. 1, 2022, 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," *International Journal of Creative Research Thoughts*, vol. 11, pp. 2320–2882, 2023, Accessed: Aug. 26, 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: https://doi.org/10.3390/atmos13020180.

[7] A H M Jakaria, M. Hossain, and Mohammad Ashiqur Rahman, "Smart Weather Forecasting Using Machine Learning:A Case Study in Tennessee," *arXiv (Cornell University)*, Aug. 2020, doi: https://doi.org/10.48550/arxiv.2008.10789.

[8] P. Kashyap and R. Venkatesh Nayak, "SMART WEATHER PREDICTION TECHNIQUES USING MACHINE LEARNING." Accessed: Nov. 21, 2024. [Online]. Available:https://www.irjmets.com/uploadedfiles/paper/volume2/issue_7_july_2020/2600/1628083097.pdf

[9] Aditya Grover, Ashish Kapoor, and Eric Horvitz. 2015. A deep hybrid model for weather forecasting. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 379–386.

[10] Mark Holmstrom, Dylan Liu, and Christopher Vo. 2016. Machine Learning Applied to Weather Forecasting.

[11] Vladimir M Krasnopolsky and Michael S Fox-Rabinovitz. 2006. Complex hybrid models combining deterministic and machine learning components for numerical climate modeling and weather prediction. Neural Networks 19, 2 (2006), 122–134.

[12] Federico Montori, Luca Bedogni, and Luciano Bononi. 2017. A Collaborative Internet of Things Architecture for Smart Cities and Environmental Monitoring. IEEE Internet of Things Journal (2017).

[13] Y Radhika and M Shashi. 2009. Atmospheric temperature prediction using support vector machines. International journal of computer theory and engineering 1, 1 (2009), 55.

[14] Rohit Kumar Yadav and Ravi Khatri. 2016. A Weather Forecasting Model using the Data Mining Technique. International Journal of Computer Applications 139

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

ORIGINALITY REPORT

7%	6%	2%	4%
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	V. Sharmila, S. Kannadhasan, A. Rajiv Kannan, P. Sivakumar, V. Vennila. "Challenges in Information, Communication and Computing Technology", CRC Press, 2024 Publication	1%
4	www.nature.com Internet Source	<1%
5	www.mdpi.com Internet Source	<1%
6	www.onlinescientificresearch.com Internet Source	<1%
7	Mehwish Kaleem, Hassan Raza, Sumaira Ashraf, António Martins Almeida, Luiz Pinto Machado. "Does ESG Predict Business Failure	<1%

in Brazil? An Application of Machine Learning Techniques", Risks, 2024
Publication

8	cse.umn.edu Internet Source	<1 %
9	www.ircobi.org Internet Source	<1 %
10	Submitted to University of Hull Student Paper	<1 %
11	ebin.pub Internet Source	<1 %
12	journals.najah.edu Internet Source	<1 %
13	Submitted to University of Canberra Student Paper	<1 %
14	soulpageit.com Internet Source	<1 %
15	Submitted to SP Jain School of Global Management Student Paper	<1 %
16	Submitted to University College London Student Paper	<1 %
17	Submitted to Pacific International Hotel Management School Student Paper	<1 %

18	quarxiv.authorea.com Internet Source	<1 %
19	www.ros-test.hw.ac.uk Internet Source	<1 %
20	Submitted to University of Bolton Student Paper	<1 %
21	S.I. Abba, Sagir Jibrin Kawu, Hamza Sabo Maccido, S.M. Lawan, Gafai Najashi, Abdullahi Yusuf Sada. "Short-term load demand forecasting using nonlinear dynamic grey-black-box and kernel optimization models: a new generation learning algorithm", 2021 1st International Conference on Multidisciplinary Engineering and Applied Science (ICMEAS), 2021 Publication	<1 %
22	d1no007.github.io Internet Source	<1 %
23	www.hydrol-earth-syst-sci.net Internet Source	<1 %
24	irjaeh.com Internet Source	<1 %
25	medium.com Internet Source	<1 %
26	mobt3ath.com Internet Source	<1 %

27	pdfcoffee.com Internet Source	<1 %
28	www.cs.odu.edu Internet Source	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		