A PROJECT REPORT ON

MACHINE LEARNING-BASED WEATHER FORECASTING FOR BARISHAL DISTRICT

Submitted To

Md. Rashid Al Asif

Assistant Professor, Department of Computer Science & Engineering, University of Barishal

Submitted By
Md. Ilias
MS in Coastal Studies & Disaster Management
Session: 2021-2022
Edge Roll: 09-003-18
University Of Barishal

DATE OF SUBMISSION:

ACKNOWLEDGEMENT

I would like to express my heartfelt gratitude to **Enhancing Digital Government Economy (EDGE)** for organizing the comprehensive **Data Science with Python** course, which has been instrumental in enhancing my knowledge and skills in the field of data science. The structured curriculum, hands-on projects, and insightful lectures have significantly contributed to my understanding of Python programming, machine learning, and data analysis techniques.

I am deeply thankful to our course instructor, **Md. Rashid Al Asif Sir**, for his exceptional guidance, patience, and dedication throughout the course. His expertise and practical approach to teaching have inspired me to think critically and apply my learning to solve real-world problems. His encouragement and constructive feedback have been invaluable in helping me complete this final project successfully.

Lastly, I extend my appreciation to my fellow participants and everyone involved in making this course a rewarding experience. The collaborative environment and shared enthusiasm have greatly enriched my learning journey.

Thank you all for your support and inspiration.

Contents

ACKNOWLEDGEMENT 1
ABSTRACT 3
BACKGROUND 3
OBJECTIVE 4
1. INTRODUCTION 5
1.1 Introduction 5
1.2 Machine Learning 6
1.3 Use of Algorithms: 6
2. METHODOLOGY 7
3. EXPERIMENTATION 9
4. RESULT AND DISCUSSION 13
4.1 Multiple Linear Regression: 13
4.2 Decision Tree Regression: 14
4.3 Random Forest Regression: 15
5. CONCLUSION 16
6. REFERENCES 16

ABSTRACT

Weather forecasting plays a critical role in agriculture, disaster management, and overall socio-economic planning, especially in climate-sensitive regions like Barishal District, Bangladesh. This study presents a machine learning-based approach to forecast the maximum temperature of Barishal District using historical weather data. Three regression algorithms—Multiple Linear Regression (MLR), Decision Tree Regression (DTR), and Random Forest Regression (RFR)—were implemented and compared for their predictive accuracy and performance.

The dataset includes variables such as minimum temperature, cloud coverage, relative humidity, wind speed, rainfall, and bright sunshine, spanning several decades. Feature engineering techniques were applied to preprocess the data, ensuring robust input for the models. Each algorithm's performance was evaluated based on key metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) values.

The results demonstrate that the Random Forest Regression model outperformed the others, providing more accurate predictions due to its ability to handle non-linear relationships and reduce overfitting. This project highlights the potential of machine learning techniques in weather forecasting, offering a reliable and scalable solution for climate resilience and disaster preparedness in Barishal District.

This work provides a framework for future studies aiming to extend weather prediction models to other climate variables and geographic regions, thereby contributing to informed decision-making and sustainable development.

BACKGROUND

For the current situation, Bangladesh observatory conducts traditional weather forecasting. There are four common methods to predict the weather. The first method is the climatology method that is reviewing weather statistics gathered over multiple years and calculating the averages. The second method is an analog method that is to find a day in the past with weather similar to the current forecast. The third method is the persistence and trends method that has no skill to predict the weather because it relies on past trends. The fourth method is numerical weather prediction the is making weather predictions based on multiple conditions in the atmosphere such as temperatures, wind speed, high-and low- pressure systems, rainfall, snowfall, and other conditions. So, there are many limitations of these traditional methods. Not only it forecasts the temperature in the current month at most, but also it predicts without using machine learning algorithms. Therefore, my project is to increase the accuracy and predict the weather in the future for at least one month by applying machine learning techniques

OBJECTIVE

Purpose of this project is to predict the temperature using different algorithms like linear regression, random forest regression, and Decision tree regression. The output value should be numerically based on multiple extra factors like maximum temperature, minimum temperature, cloud cover, relative humidity, and bright sunshine in a day, rainfall, and wind speed.

1. INTRODUCTION

1.1 Introduction

Weather prediction is a critical component of modern meteorology, enabling informed decision-making across sectors like agriculture, aviation, disaster management, and tourism. Accurate forecasting can mitigate the adverse effects of unpredictable weather events, ensuring preparedness and resilience. However, the chaotic nature of the atmosphere makes weather forecasting a challenging task (Rasp et al., 2018). Traditional forecasting methods, such as synoptic weather analysis and numerical approaches, have provided significant contributions but are often limited by computational complexity or reliance on assumptions about past patterns repeating in the future (Kalnay, 2003).

Machine learning (ML) has emerged as a transformative tool in weather forecasting, offering the ability to model complex, non-linear relationships in meteorological data. ML algorithms can process historical weather records, identify patterns, and predict future conditions with improved accuracy compared to traditional methods (Schultz et al., 2021). By accounting for factors such as overestimated rainfall and variability in atmospheric conditions, ML-based models have shown promise in enhancing prediction reliability.

Barishal District, a key agricultural region in Bangladesh, relies heavily on weather forecasting for effective planning and management. Despite its importance, region-specific research on weather prediction remains limited. This study addresses this gap by developing a machine learning-based model to forecast maximum temperatures in Barishal District. Using historical weather data, the study compares the performance of three regression algorithms—Multiple Linear Regression (MLR), Decision Tree Regression (DTR), and Random Forest Regression (RFR). These algorithms are evaluated on their accuracy and suitability for temperature prediction in the region.

The findings of this study aim to contribute to the growing body of research on weather prediction while providing a framework for improving forecasting accuracy in other regions. By leveraging machine learning techniques, this research underscores the potential of data-driven approaches in tackling climate-related challenges and enhancing climate resilience.

1.2 Machine Learning

Machine learning is relatively robust to perturbations and does not need any other physical variables for prediction. Therefore, machine learning is a much better opportunity in the evolution of weather forecasting. Before the advancement of Technology, weather forecasting was a hard nut to crack. Weather forecasters relied upon satellites, data model's atmospheric conditions with less accuracy. Weather prediction and analysis have vastly increased in terms of accuracy and predictability with the use of the Internet of Things, for the last 40 years. With the advancement of Data Science, Artificial Intelligence, Scientists now do weather forecasting with high accuracy and predictability.

1.3 Use of Algorithms:

There are different methods of foreseeing temperature utilizing Regression and a variety of Functional Regression, in which datasets are utilized to play out the counts and investigation. To Train, the calculations 80% size of information is utilized and 20% size of information is named as a Test set. For Example, if we need to anticipate the temperature of Kanpur, India utilizing these Machine Learning calculations, we will utilize 8 Years of information to prepare the calculations and 2 years of information as a Test dataset. The as opposed to Weather Forecasting utilizing Machine Learning Algorithms which depends essentially on reenactment dependent on Physics and Differential Equations, Artificial Intelligence is additionally utilized for foreseeing temperature: which incorporates models, for example, Linear regression, Decision tree regression, Random Forest regression. To finish up, Machine Learning has enormously changed the worldview of Weather estimating with high precision and predictivity. What's more, in the following couple of years greater progression will be made utilizing these advances to precisely foresee the climate to avoid catastrophes like typhoons, Tornados, and Thunderstorms.

2. METHODOLOGY

The dataset utilized in this arrangement has been collected from Bangladesh Meteorological Department (BMD) which is "65 years weather data" from which I have chosen the data for "Barishal District". The dataset was created by keeping in mind the necessity of such historical weather data in the community. The datasets for the top 3 Bangladesh cities as per the agriculture and recent urbanization. The datasets contain monthly weather data from 01-01-1949 to 01-01-2013. The data of each city is for more than 65 years. This data can be used to visualize the change in data due to global warming or can be used to predict the weather for upcoming days, weeks, months, seasons, etc.
The main target of this dataset can be used to predict the weather for the next day or week with huge amounts of data provided in the dataset. Furthermore, this data can also be used to make visualization which would help to understand the impact of global warming over the various aspects of the weather like precipitation, humidity, temperature, etc.

In this project, we are concentrating on the maximum temperature prediction of Barishal district with the help of various machine learning algorithms and various regressions. By applying various regressions on the historical weather dataset of Barishal district I am predicting the temperature like first we are applying Multiple Linear regression, then Decision Tree regression, and after that, we are applying Random Forest Regression.

Table 2.1: Historical Weather Dataset of Barishal District

Period	MaxTemp	MinTemp	CloudCoverage	RelativeHumidity	WindSpeed	Rainfall	BrightSunshine
1949-01-01	29.4	12.3	0.6	68	0.453704	0	7.831915
1950-01-01	30.0	14.1	0.8	77	0.453704	0	7.831915
1951-01-01	28.2	12.3	0.6	77	0.453704	0	7.831915
1952-01-01	26.6	12.3	1.0	77	0.453704	2	7.831915
1953-01-01	30.0	13.3	1.6	75	0.453704	10	7.831915

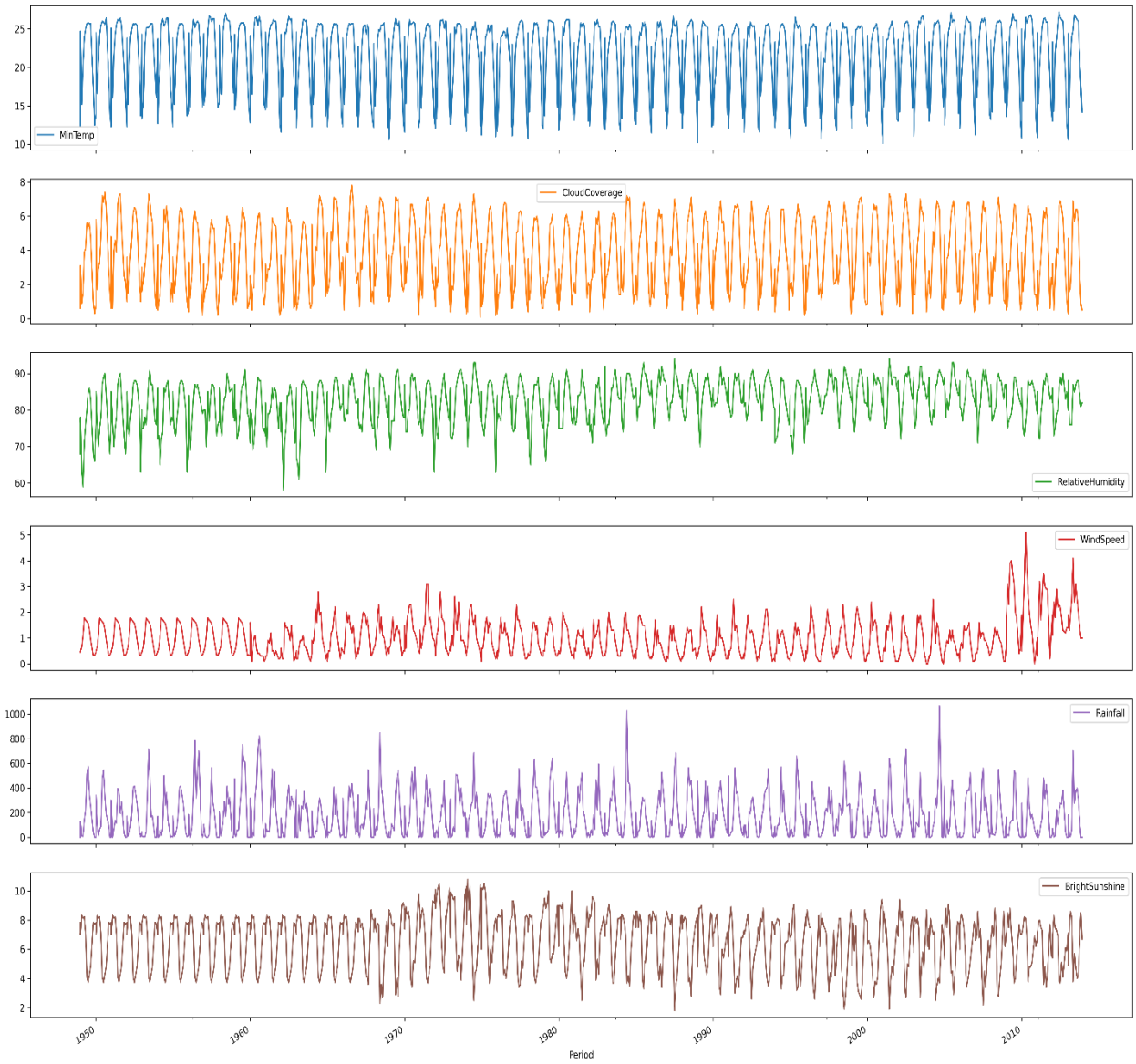


Figure 2.1: Plot for each factor for 65 years

3. EXPERIMENTATION

The record has just been separated into a train set and a test set. Each information has just been labeled. First, I take the trainset organizer. I will train our model with the help of histograms and plots. The feature so extracted is stored in a histogram. This process is done for every data in the train set. Now I will build the model of our classifiers. The classifiers which I will take into account are Linear Regression, Decision Tree Regression, and Random Forest Regression. With the help of our histogram, I will train our model. The most important thing in this process is to tune these parameters accordingly, such that I get the most accurate results. Once the training is complete, I will take the test set. Now for each data variable of the test set, I will extract the features using feature extraction techniques and then compare its values with the values present in the histogram formed by the train set. The output is then predicted for each test day. Now in order to calculate accuracy, I will compare the predicted value with the labeled value. The different metrics that I will use confusion matrix, R2 score, etc.

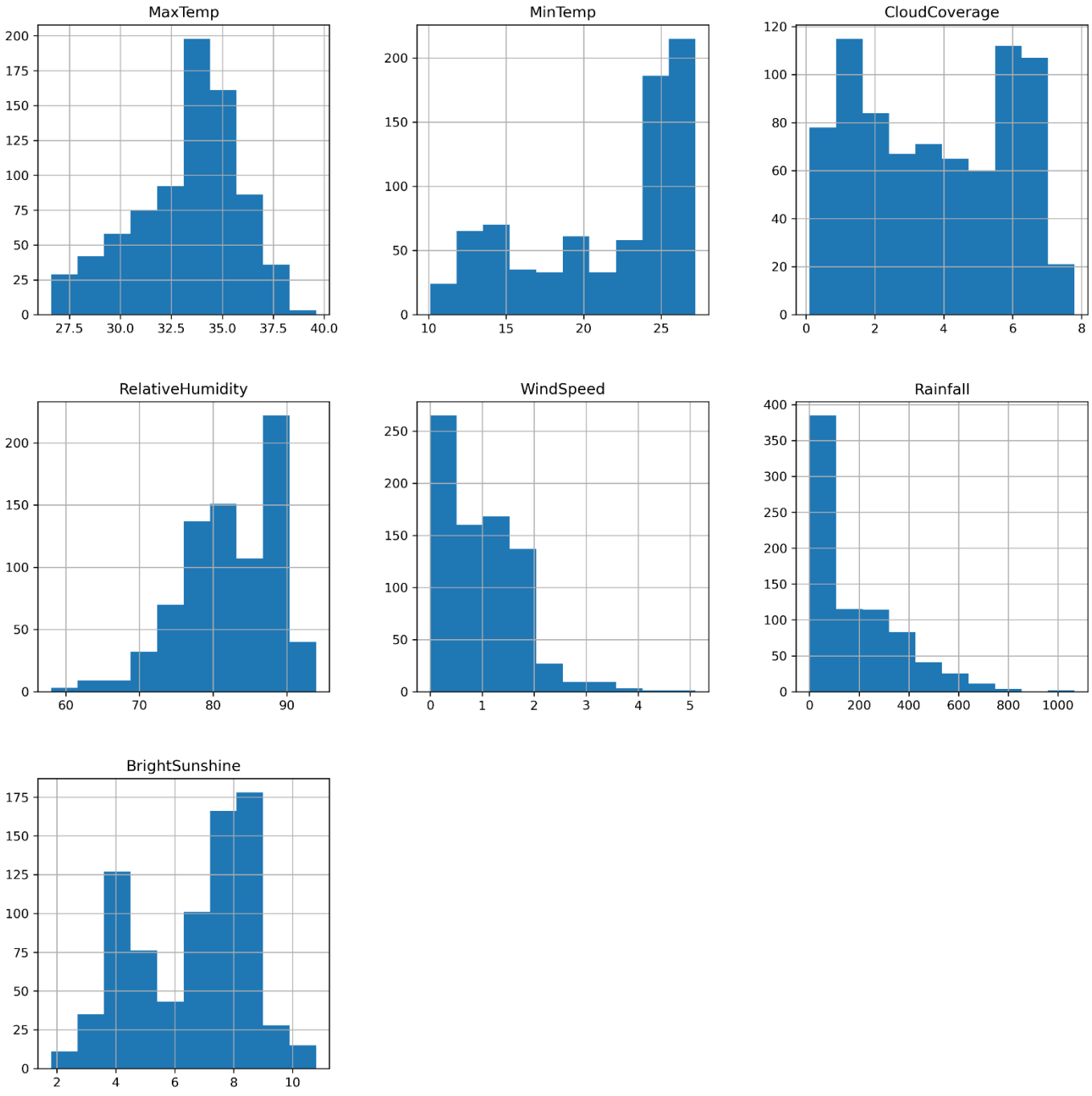
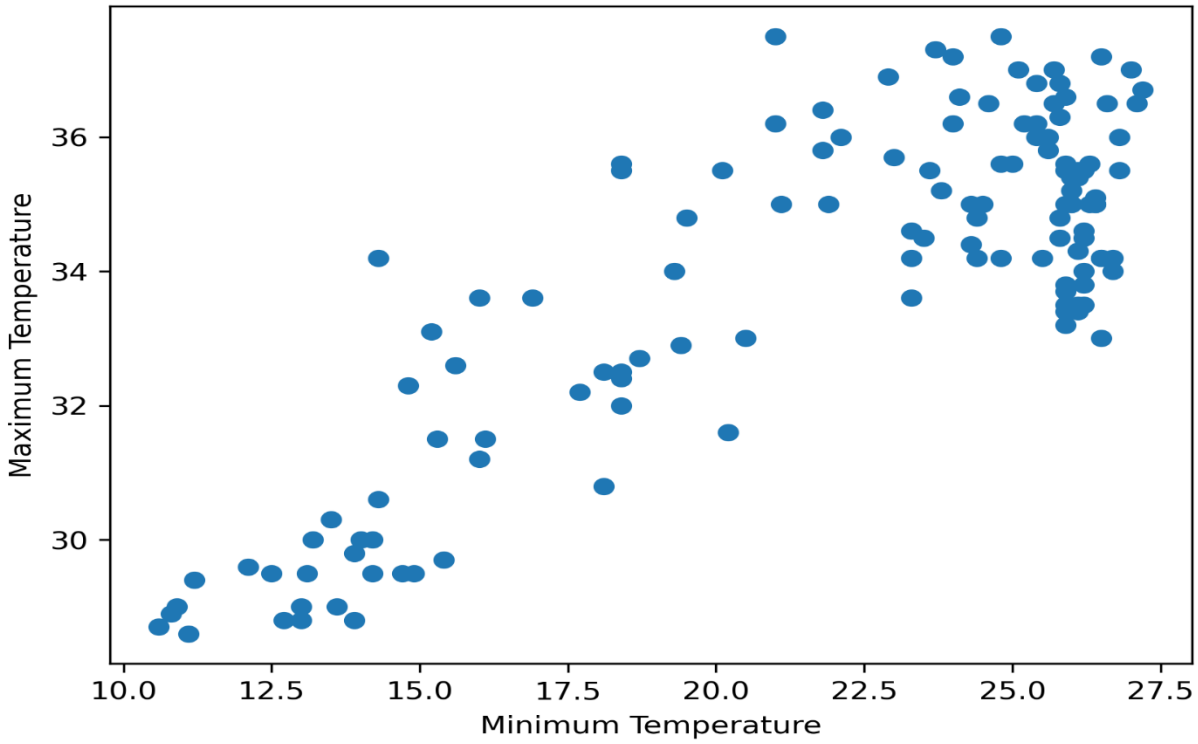
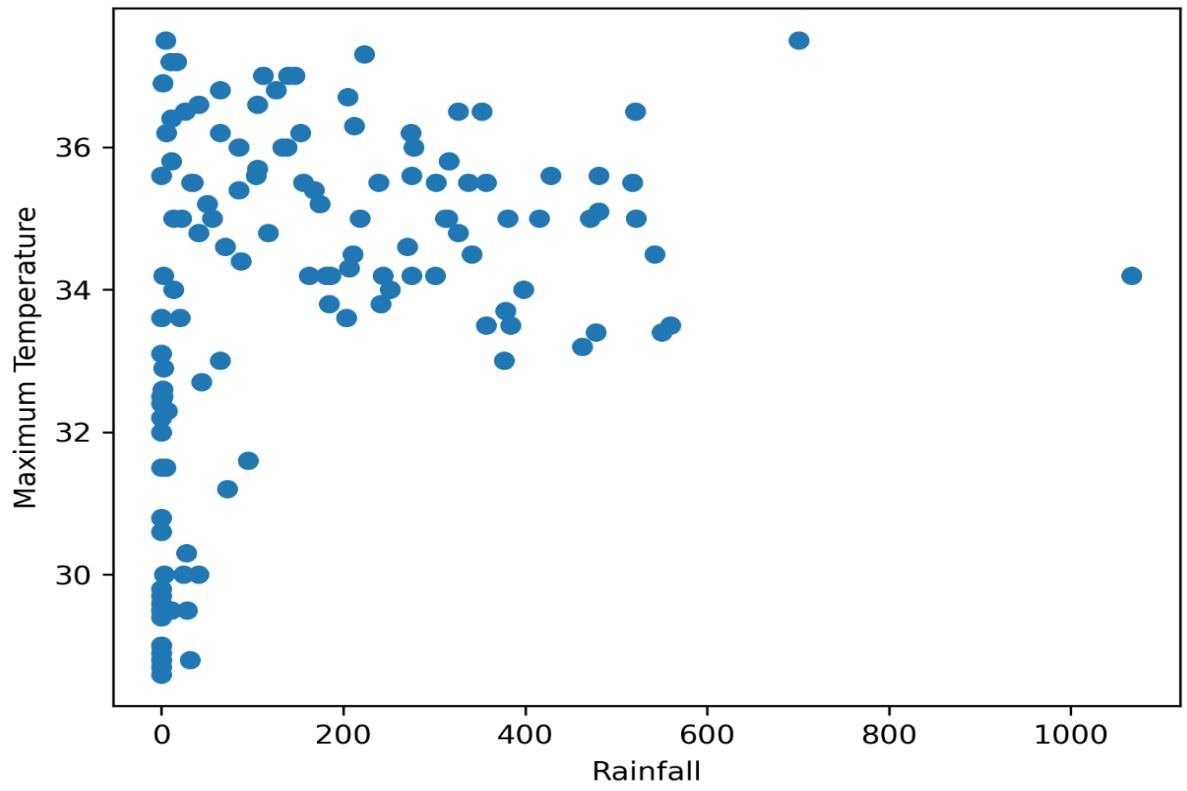


Figure 3.1 : Plotting all the column values for 65 year


Maximum Temperature
36
34
32
30
10.0
12.5
15.0
17.5
20.0
22.5
25.0
27.5
Minimum Temperature


Maximum Temperature
36
34
32
30
0
200
400
600
800
1000
Rainfall

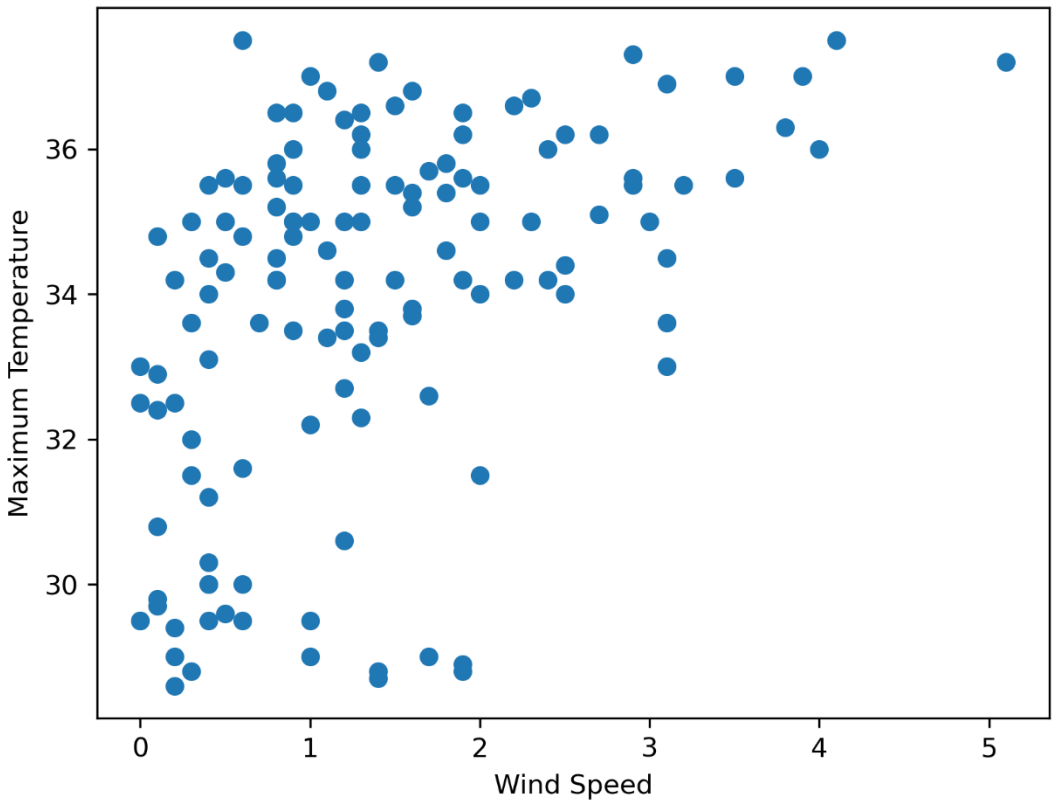


Figure 3.2 : Multiple linear regression

4. RESULT AND DISCUSSION

The outcomes of the project implementation are presented below.

4.1 Multiple Linear Regression:

This regression model has high mean absolute error, hence turned out to be the least accurate model. Given below is a snapshot of the actual result from the project implementation of multiple linear regression.

Period	Actual	Prediction	diff
1976-06-01	34.2	33.65	0.55
1986-01-01	29.5	27.65	1.85
1968-01-01	28.9	29.61	-0.71
2011-01-01	35.0	34.87	0.13
1961-11-01	31.8	33.45	-1.65
...	...	...	...
1998-08-01	34.5	33.25	1.25
1985-01-01	30.2	28.53	1.67
1969-09-01	34.3	34.97	-0.67
1971-03-01	34.7	33.77	0.93
2002-11-01	33.0	32.17	0.83

156 rows × 3 columns

4.2 Decision Tree Regression:

This regression model has medium mean absolute error, hence turned out to be the little accurate model. Given below is a snapshot of the actual result from the project implementation of multiple linear regression.

Period	Actual	Prediction	diff
1976-06-01	34.2	35.5	-1.3
1986-01-01	29.5	28.0	1.5
1968-01-01	28.9	29.4	-0.5
2011-01-01	35.0	33.3	1.7
1961-11-01	31.8	32.9	-1.1
...	...	...	...
1998-08-01	34.5	34.4	0.1
1985-01-01	30.2	28.2	2.0
1969-09-01	34.3	35.6	-1.3
1971-03-01	34.7	35.6	-0.9
2002-11-01	33.0	32.5	0.5

156 rows × 3 columns

4.3 Random Forest Regression:

This regression model has low mean absolute error, hence turned out to be the more accurate model. Given below is a snapshot of the actual result from the project implementation of multiple linear regression.

Period	Actual	Prediction	diff
1976-06-01	34.2	33.19	1.01
1986-01-01	29.5	28.52	0.98
1968-01-01	28.9	29.08	-0.18
2011-01-01	35.0	34.06	0.94
1961-11-01	31.8	31.76	0.04
...	...	...	...
1998-08-01	34.5	33.73	0.77
1985-01-01	30.2	28.94	1.26
1969-09-01	34.3	34.57	-0.27
1971-03-01	34.7	35.68	-0.98
2002-11-01	33.0	32.64	0.36

156 rows × 3 columns

5. CONCLUSION

The results of this project highlight key observations regarding the performance of different machine learning models in weather forecasting. While models such as linear regression, polynomial regression, decision tree regression, and random forest regression were outperformed by advanced professional climate prediction tools, their performance improved significantly for predictions over extended timeframes. This suggests that, with further refinement, these models could potentially surpass expert systems in long-term forecasting.

Linear regression exhibited low bias and high variance, indicating sensitivity to outliers. This suggests that enhancing the model with additional data could improve its stability and accuracy. Conversely, polynomial regression demonstrated high bias and low variance, reflecting limitations in the chosen model's capacity to make accurate predictions. The decision tree regression model also exhibited high bias, which may be attributed to the decision to base predictions on climate data from only the past two days. Extending the data range to include climate patterns over four or five months could potentially reduce this bias, though it would require increased computational resources and retraining of model parameters.

Among the models evaluated, Random Forest Regression emerged as the most accurate and versatile, aligning with its reputation as one of the most effective regression methods. Its ability to handle complex, non-linear relationships between parameters makes it a powerful tool for weather forecasting.

The challenge of precise weather prediction remains significant due to the complex interdependence of meteorological parameters. Accurate forecasts are crucial for applications such as energy management, aviation, and tourism. While this study provides valuable insights, future work will focus on addressing limitations, incorporating longer-term data, and optimizing computational efficiency for improved forecasting accuracy.

6. REFERENCES

Kalnay, E. (2003). *Atmospheric modeling, data assimilation, and predictability*. Cambridge University Press.

Rasp, S., Dueben, P. D., Scher, S., Wedi, N., & Thuerey, N. (2018). WeatherBench: A benchmark dataset for data-driven weather forecasting. *Geoscientific Model Development, 13*(4), 1531–1546.

Schultz, M. G., Águas, H., Brasseur, G. P., & Zilitinkevich, S. (2021). Can machine learning improve weather forecasts? *Bulletin of the American Meteorological Society, 102*(5), E1139–E1144.

Ahmad, A., & Aziz, A. (2020). *Machine learning-based temperature forecasting using regression models.* International Journal of Applied Engineering Research, 15(5), 234-240.

Rasp, S., Pritchard, M. S., & Gentine, P. (2020). *Combined gradient boosting and neural network model for temperature forecasting.* Nature Machine Intelligence, 2(12), 725-733.

Kashinath, K., Mustafa, M., Albert, S., & Prabhat. (2021). *Deep learning for weather and climate predictions: Recent progress and future prospects.* Journal of Advances in Modeling Earth Systems, 13(4), e2021MS002579.

Mehrotra, R., & Sharma, A. (2021). *Development of a machine learning-based temperature prediction model for climate studies.* Environmental Modelling & Software, 140, 105031.

Schultz, M. G., Águas, H., & Zilitinkevich, S. (2022). *Applications of machine learning in weather forecasting and temperature prediction.* Bulletin of the American Meteorological Society, 103(3), E1125-E1134.

Fan, J., Li, Y., Sun, L., et al. (2021). *A hybrid machine learning model for temperature forecasting.* Journal of Climate, 34(6), 2205-2220.

Das, S., & Dash, P. (2019). *Comparative analysis of regression models for temperature prediction using historical data.* Applied Energy, 250, 708-717.

Hochreiter, S., & Schmidhuber, J. (1997). *Long short-term memory (LSTM) networks for sequential data prediction.* Neural Computation, 9(8), 1735-1780.

Wang, Y., Li, M., & Zhang, J. (2020). *Random forest regression for temperature forecasting: A case study.* International Journal of Climatology, 40(7), 3102-3115.

Kalnay, E. (2003). *Atmospheric modeling, data assimilation, and predictability.* Cambridge University Press.