Weather Forecasting Using Streamlit

N. M. SUHAIL[1], K. VANITHASRI[2], P. SANJEEV KUMAR[3], K. SAKTHIVEL[4], N. THAMEEMULLAH[5]

[1, 3, 4, 5]*Department of computer science and Engineering, Aalim Muhammed Salegh College of Engineering (Affiliated to Anna University), Chennai, Tamil Nadu, India.*

[2]*Assistant Professor, Department of Computer Science and Engineering, Aalim Muhammed Salegh College of Engineering (Affiliated to Anna University), Chennai, Tamil Nadu, India.*

***Abstract—* Weather prediction is a critical aspect of our daily lives, impacting decisions ranging from what to wear to how we plan our outdoor activities. In this project, we aim to leverage machine learning techniques to predict weather temperatures based on various meteorological parameters. The primary objective is to develop accurate and reliable temperature forecasting models that can assist individuals and organizations in making informed decisions. This Weather Temperature Prediction project represents a valuable contribution to the field of weather forecasting and demonstrates the potential of machine learning in solving real-world problems. Our results demonstrate that machine learning models can provide accurate temperature predictions, with the Random Forest Regression model consistently outperforming the others. By deploying our model via Streamlit, we have created a user-friendly tool that empowers individuals and organizations to make data-driven decisions based on weather forecasts. This project not only showcases the application of machine learning in meteorology but also highlights the practicality of deploying such models in real-world scenarios. As weather plays a crucial role in various industries and daily activities, accurate temperature predictions can significantly benefit society.**

***Index Terms-* Weather Forecast, Machine Learning, Python, Streamlit, Temperature Prediction, ML models**

I. INTRODUCTION

Weather is an integral part of our daily lives, influencing everything from what we wear to how we plan our outdoor activities. Accurate temperature forecasting is crucial for making informed decisions, whether you're preparing for a weekend hike or managing energy consumption in a city. This Weather Temperature Prediction project aims to harness the power of machine learning to provide reliable and precise temperature predictions based on a range of meteorological factors.

In this project, we collected historical weather data from a trusted source and employed advanced data preprocessing techniques to ensure data quality. We then explored various machine learning algorithms, including Multiple Linear Regression, Decision Tree Regression, and Random Forest Regression, to model temperature patterns. These models were evaluated using key metrics such as Mean Absolute Error (MAE) and R-squared (R2) scores to gauge their accuracy and reliability.

This Weather Temperature Prediction project is not just a demonstration of machine learning in meteorology but also a practical solution to a real-world problem. Weather impacts various sectors, including agriculture, transportation, and energy management, and accurate temperature forecasts are critical for success in these fields.

II. OVERVIEW OF IMPLEMENTED SYSTEM

In the proposed system, an application has been developed using streamlit library. This application gets the real-time data from humidity, heat index, cloud cover, sun hours and pressure to predict the accurate temperature on the present day. Fig. 1 demonstrates the block-level diagram of the system. The weather data is preprocessed and used to train the machine learning mode. Regression algorithms are used to predict the temperature. The prediction result from the model is used to find the performance of each model Then the best model is found by performance analysis and that model is used for prediction.

The application is made using streamlit. The core concept of Streamlit revolves around the idea of turning a Python script into a web app. Users can write a script that generates plots, tables, and other interactive elements using libraries like Pandas, Matplotlib, or Plotly. They can then use Streamlit to easily add web functionalities to this script.

The forecast.ipynb module includes the machine learning model which is trained according to the Multiple Linear Regression, Random forest Regression, Decision Tree Regression. It collects the data from the dataset kanpur.csv to train the ML models. The data include minimum temperature, maximum temperature, cloud cover, humidity, sun hours, heat index, precipitation, windspeed and pressure. The app.py module uses the streamlit library to create the application including the backend work and GUI.

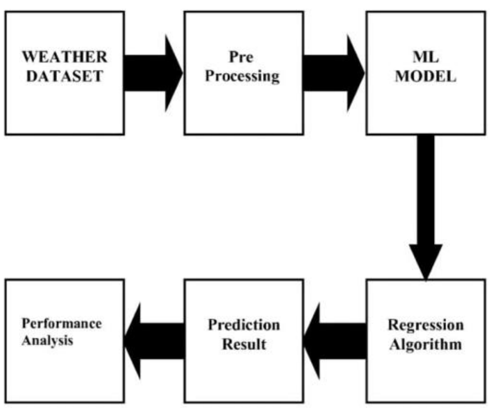


Fig. 1. Block diagram of the system

The snap shot of the developed GUI is shown in Fig. 2.

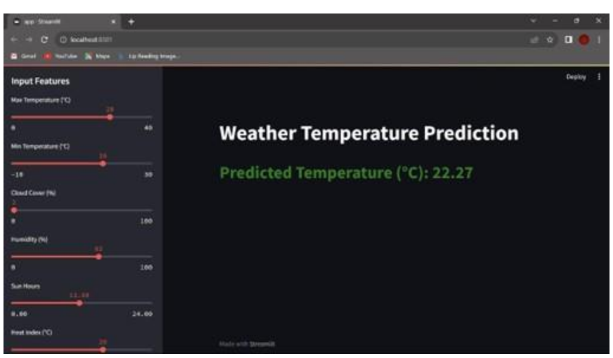


Fig. 2. GUI of the developed application.

III. MACHINE LEARNING APPROACH

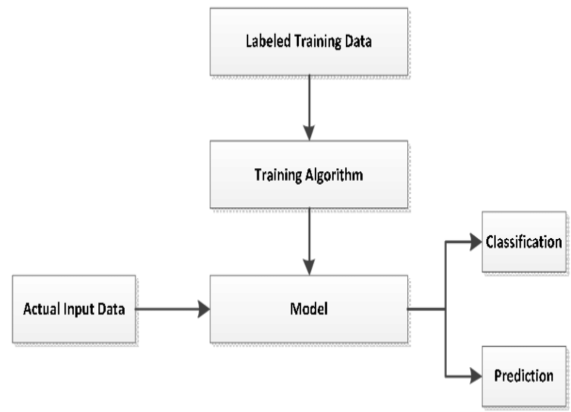


Fig. 3. Machine learning-based prediction system.

Several steps are followed for implementing the machine learning approach in this work. The steps are as follows:

A. Data Collection and preprocessing

We collected historical weather data from a reliable source and pre-processed it to ensure data quality. It is Kanpur's weather data for 10 years from 2009 to 2019. This involved handling missing values, selecting relevant features, and transforming the dataset into a suitable format for modelling.

B. Data Modeling

We explored multiple machine learning algorithms for temperature prediction. This included implementing a Multiple Linear Regression model to establish baseline performance. We also experimented with Decision Tree Regression and Random Forest Regression models to capture nonlinear relationships within the data.

C. Model Buildng

To assess the accuracy of our models, we used various performance metrics such as Mean Absolute Error (MAE), Mean Squared Error(MSE), and R-squared (R2) scores. This evaluation process helped us select the most suitable model for temperature forecasting.

IV. MULTIPLE LINEAR REGRESSION

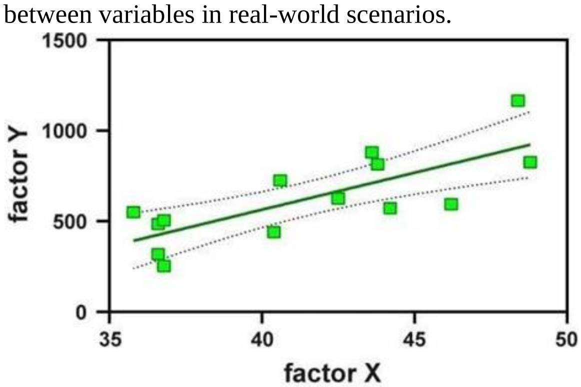
The goal of this method is to find the best-fitting linear equation that describes the relationship between the independent variables and the dependent variable.

In mathematical terms, the model can be represented as:
[Y = beta_0 + beta_1X_1 + beta_2X_2 + ... + beta_nX_n + varepsilon]

The aim is to estimate the coefficients in such a way that the sum of squared differences between the observed and predicted values (the residual sum of squares) is minimized. This is typically done using a technique called least squares, which minimizes the sum of the squared vertical distances between the observed data points and the regression line.

Assumptions play a crucial role in multiple linear regression. These include the linearity assumption (the relationship between variables is linear), independence of errors (the errors are not correlated with each other), normality of errors (the errors follow a normal distribution), and homoscedasticity (constant variance of errors).

Multiple Linear Regression provides a powerful tool for understanding and quantifying the relationships between variables in real-world scenarios.

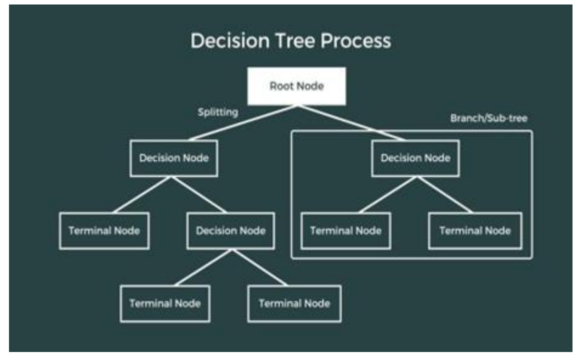


V. DECISION TREE REGRESSION

Decision tree regression is a machine learning algorithm used for both classification and regression tasks.

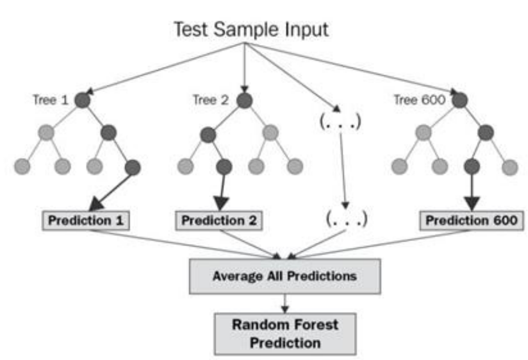
During the training process, the algorithm seeks to find the best splits at each node. This is done by evaluating different features and their respective thresholds, aiming to minimize the variance of the target variable within each subset. The goal is to create partitions that result in homogeneous groups with respect to the dependent variable.

Once the tree is constructed, making predictions is straightforward. Starting from the root node, the algorithm traverses the tree based on the featurevalues of the data point in question, ultimately reaching a leaf node. The predicted value is then determined by the average of the target values in that leaf node's training samples.



VI. RANDOM FOREST REGRESSION

Random Forest regression is a powerful machine learning algorithm that combines the principles of ensemble learning with decision tree regression. It works by constructing a multitude of decision trees during the training phase. Each tree is built using a random subset of the training data and a random subset of the features. This randomness injects diversity into the individual trees, which helps to reduce overfitting and improve predictive accuracy.



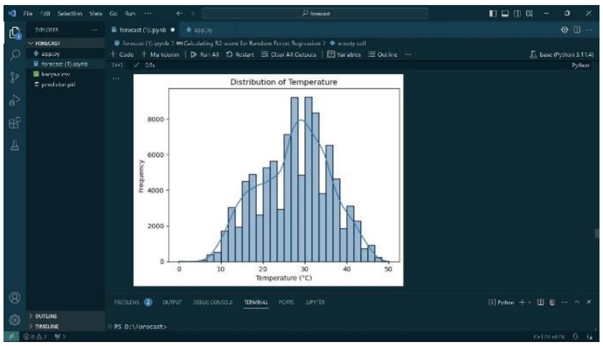
D. Deployment with Streamlit

To make our temperature prediction accessible to a broader audience, we deployed our best-performing model using Streamlit, a user-friendly web application framework. This Streamlit application allows users to input relevant weather features and receive real-time temperature predictions.

VII. DATA VISUALIZATION

A. Distribution of Temperature

The histogram illustrates the temperature (tempC) distribution, showing a bell-shaped pattern with a central peak, signifying a common temperature range.



B. Temperature Distribution by Month

The boxplot demonstrates the temperature distribution across months. It reveals seasonal variations, with higher temperatures during the summer months and lower temperatures in winter.

VIII. RESULTS

The bar plots illustrate the model performance in terms of these metrics.

For MAE and MSE, lower values indicate better performance, and the Random Forest model outperforms the others with the lowest MAE and MSE.

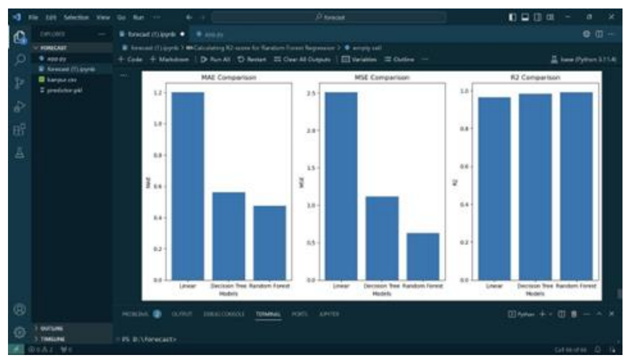
R2 measures the goodness of fit, with values closer to 1 indicating better fitting models.

The Random Forest model also exhibits the highest R2, suggesting it explains the data variance more effectively.

Overall, the Random Forest model appears to be the most suitable choicefor temperature prediction based on these metrics.

TABLE I. COMPARISON OF ALGORITM PERFORMANCE

Algorithm	MAE	MSE	R2 Score
Multiple Linear Regression	1.20	2.51	0.96
Decision Tree Regression	0.56	1.12	0.98
Random Forest Regression	0.47	0.63	0.99



CONCLUSION

In the project's conclusion, an exhaustive evaluation of various machine learning models for weather temperature prediction was undertaken. Despite promising results, these models consistently fell short in comparison to expert climate forecasting tools, especially in short-term predictions. However, a notable trend emerged as the models displayed a significant reduction in prediction errors over time, indicating the potential for them to outperform expert tools in longer-term forecasts. It was revealed that linear regressionexhibited low bias but high variance, making it sensitive to outliers and suggesting the need for additional data collection to enhance its performance. Conversely, polynomial regression demonstrated high bias and low variance. Practical regression, based on a short-term weather history, exhibited high bias and poor decision-making ability. Future work could involve extending historical data for more comprehensive insights. Notably, Random Forest Regression proved to be the most accurate model, reflecting its popularity for its versatility and high precision. The project underscores the challenges posed by the intricate nature of weather parameters and their diverse value ranges while emphasizing the crucial role of accurate weather predictions in critical systems like power facilities, airports, and tourism centers.

REFERENCES

[1] M. M. Ali, P. S. V. Jagadeesh, I.-I. Lin, and Je-Yuan Hsu, "A Neural Network Approach to Estimate Tropical Cyclone Heat Potential in the Indian Ocean" IEEE Geosciences and Remote Sensing Letters, Vol. 9, No. 6, Nov 2012, pp 1114-1117.

[2] Diganta Kumar Sarma, Mahen Konwar, Jyotirmoy Das, Srimanta Pal, and Sanjay Sharma, "A Soft Computing Approach for Rainfall Retrieval From the TRMM Microwave Imager", IEEE Transactions on Geosciences and Remote Sensing, Vol. 43, No. 12, December 2005, pp 2879-2885.

[3] Sutapa Chaudhuri, Surajit Chattopadhyay, "Neurocomputing based shortrange prediction of some meteorological parameters during the pre-monsoon season", Springer-Verlag 2005, Journal of Soft Computing (2005) Vol 9, pp 349–354.

[4] F. Mekanik and M. A. Imteaz, "A Multivariate Artificial Neural Network Approach for Rainfall Forecasting: Case Study of Victoria", Australia, Proceedings of the World Congress on Engineering and Computer Science 2012,Vol I WCECS 2012, October 24- 26, 2012, San Francisco, USA.

[5] Kin C. Luk, J. E. Ball and A. Sharma, "An Application of Artificial Neural Networks for Rainfall Forecasting", Mathematical and Computer Modelling, Volume 33, pp 683-693, 2001.

[6] Mohsen Hayati and Zahra Mohebi, "Temperature Forecasting Based on Neural Network Approach", World Applied Sciences JournalVolume 2, 2007.

[7] Ahmad, Drlatief & Biswas, Asim & Warland, Jon & Anjum, Insha. (2023). Weather Forecasting. 10.1007/978-981-99-4863-5_7.

[8] More, Nayana. (2024). Weather Forecasting: A Comprehensive Exploration of IoT and Machine Learning. INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT. 08. 1-5. 10.55041/IJSREM31952

[9] Raghavendra, Sujay. (2022). Introduction to Streamlit. 10.1007/978-1-4842-8983-9_1.

[10] Gomathy, C K. (2022). WEATHER FORECASTING APPLICATION USING PYTHON

[11] Veeramsetty, Venkataramana & Kiran, Prabhu & Sushma, Munjampally & Salkuti, Surender Reddy. (2023). Weather Forecasting Using Radial Basis Function Neural Network in Warangal, India. Urban Science. 7. 68. 10.3390/urbansci7030068.

[12] Jani, Nidhi & Gupta, Rajeev & Bharti, Drsantosh & Jain, Dr. (2022). Advancements in Weather Forecasting With Deep Learning. 10.4018/978-1-6684-3981-4.ch006.

[13] Saw, Montu & Jain, Sachin & Ragavendran, U. (2023). Review Paper on Weather Forecasting App. International Journal of Advanced Research in Science, Communication and Technology. 83-89. 10.48175/IJARSCT-10726

[14] Khorasani, Mohammad & Abdou, Mohamed & Hernandez Fernandez, Javier. (2022). Building Streamlit Components. 10.1007/978-1-4842-8111-6_10.

[15] Kumar, Vimal & Kumar, Rajesh & Bhasin, Chandra Praksh & Kumar, Alesh & Scholar, Ph. (2022). Weather Forecasting through Synoptic Technique.

[16] Tariq, Muhammad Usman & Al-Hosani, Reem. (2022). The Essence of Smart Weather Forecasting Using Machine Learning.