International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 12, Issue 10, October 2023)

An Efficient Machine learning Technique for Prediction of Weather Forecasting

Abhishek Singh[1], Hitesh Gupta[2]

[1]Research Scholer, [2] Professor, Department of Computer Science & Engineering, LNCT, Bhopal (M.P), India

[1]as3496963@gmail.com, [2]hiteshg@lnct.ac.in

Abstract-- **Weather forecasting is a critical aspect of meteorology, impacting various sectors such as agriculture, transportation, and disaster management. Accurate predictions of weather conditions are essential to make informed decisions and mitigate potential risks. Traditional numerical weather prediction models have their limitations and recent advancements in machine learning techniques. This paper presents the random forest (RF) technique to improve the accuracy of prediction of weather forecasting.**

IndexTerms **– Machine Learning, Weather Forecasting, Prediction, RF.**

I. INTRODUCTION

Weather forecasting is a fundamental and indispensable part of modern society, impacting numerous facets of our daily lives, from planning outdoor activities and agricultural practices to safeguarding against natural disasters and optimizing transportation systems. The ability to predict weather conditions with precision is not only of scientific interest but also crucial for informed decision-making and risk mitigation.

Traditionally, numerical weather prediction models have been the primary tools for forecasting. These models, while valuable, have inherent limitations, particularly when it comes to capturing the intricate, non-linear relationships between various meteorological parameters. As climate patterns become increasingly complex and dynamic due to factors like climate change, there is a growing need for innovative approaches that can augment the accuracy and lead time of weather predictions.

In this context, the utilization of machine learning techniques has emerged as a promising avenue to enhance weather forecasting. One such technique is the Random Forest algorithm, an ensemble learning method known for its adaptability and robust performance. By leveraging Random Forest, we can potentially overcome the limitations of traditional forecasting methods and improve the quality and reliability of weather predictions.

Predicting forecasting the state of the atmosphere at a certain place and time is referred to as weather forecasting. It entails the examination of a number of different atmospheric factors, such as temperature, humidity, air pressure, wind speed and direction, and patterns of precipitation. In order to create these forecasts, meteorologists rely on scientific models, data collected throughout history, and observations gleaned from weather stations, satellites, and radar systems.

Models that are used for numerical weather prediction (also known as NWP) serve as the basis for weather forecasting. For the purpose of simulating how the atmosphere behaves, these models make use of intricate mathematical calculations. They are able to make an approximation of how the atmosphere will change over time by taking into account the starting circumstances and following the rules of physics. In order to create predictions, these models are processed by very powerful supercomputers. In order to effectively initialize the models, it is essential to get observations from a variety of sources, including weather stations, weather balloons, satellites, and other devices. These observations offer data on the present weather conditions, which are used as the starting point for the calculations that are performed by the models. Radars may also monitor precipitation, storms, and other severe weather phenomena in real time. This is another purpose for weather radars. Forecasting the weather has become much more accurate throughout the course of time because to the steady march of technological advancement. Because they are based on more recent data, short-range predictions, which may encompass anything from a few hours to a few days, often provide more accurate results. Due to the growing amount of uncertainty, medium-range predictions, which cover a period of several days up to a week, are less reliable. Long-range predictions, or those that stretch beyond than a week, have the lowest level of accuracy but may still provide some direction in a broad sense.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 12, Issue 10, October 2023)

The ability to accurately predict the weather is vital in a wide variety of fields, including agriculture, aviation, transportation, disaster relief, and individual day-to-day life planning. It facilitates our preparation for severe weather occurrences, enables us to make choices based on accurate information, and reduces the likelihood of possible threats.

The use of machine learning strategies to the practice of weather forecasting is a relatively new discipline that seeks to enhance both the precision and effectiveness of weather forecasting. The capabilities of traditional numerical weather prediction models to handle complicated connections and capture detailed patterns found in meteorological data are limited. On the other hand, machine learning algorithms have the ability to overcome these difficulties by learning from previous data on the weather and producing forecasts based on patterns that have been discovered.

II. Proposed Methodology

The methodology can be understand by using following flow chart-

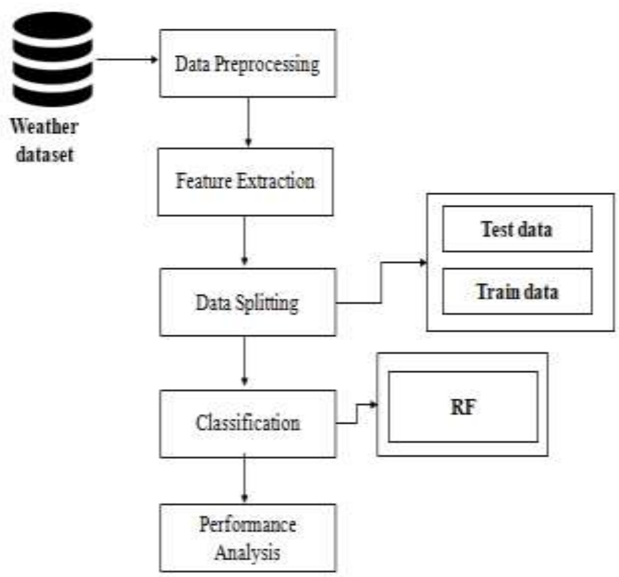


Figure 1: Flow chart

Steps-

1. Data Collection:

Gather historical weather data from reliable sources, including temperature, humidity, wind speed, atmospheric pressure, and other meteorological parameters. Ensure that the data is time-stamped, spans a sufficiently long period, and covers the geographic area of interest.

2. Data Preprocessing:

Clean the data to address missing values, outliers, and inconsistencies.

Normalize or standardize the data to bring all features to a common scale.

Divide the dataset into training, validation, and testing sets for model development and evaluation.

3. Feature Selection and Engineering:

Identify relevant weather features that are likely to influence the target variable (e.g., temperature, humidity, wind direction).

Create derived features or transformations if necessary to capture meaningful information.

4. Random Forest Model Configuration:

Choose the number of decision trees (n_estimators) in the Random Forest ensemble.

Define hyperparameters such as maximum tree depth, minimum samples per leaf, and feature selection criteria.

Consider implementing random feature subsets (random subspaces) and bootstrapping for diversity within the ensemble.

5. Model Training:

Train the Random Forest model using the training dataset.

The model will learn to make predictions based on historical weather data.

III. Simulation And Results

The implementation and simulation of the proposed work is done over python spyder 3.7 software.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 12, Issue 10, October 2023)

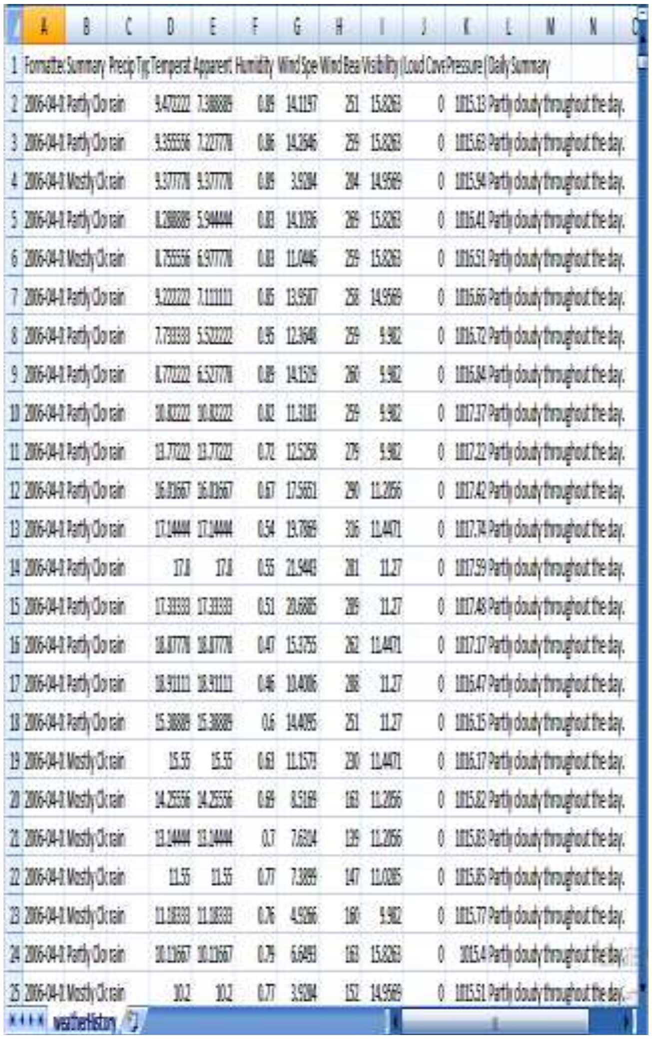


Figure 2: Original dataset in .csv file

The figure 2 is showing the dataset, which is taken from the kaggle machine learning website.

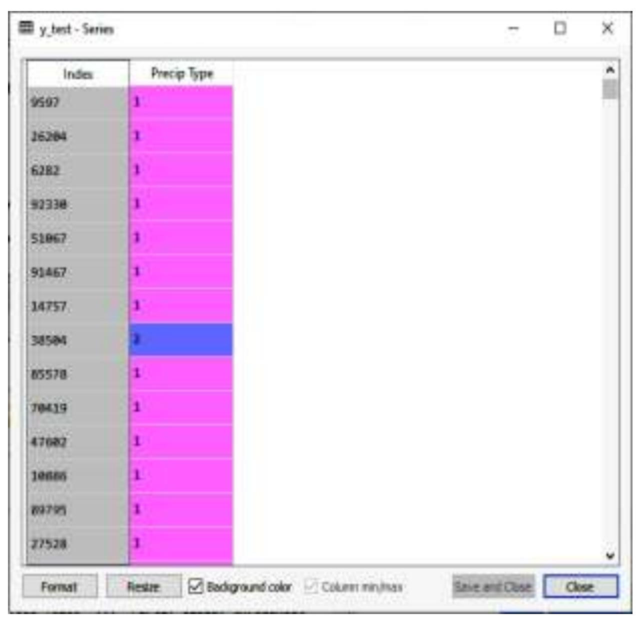
X_test - DataFrame

Index	Summary	Temperature (C)	arent Temperature	Humidity	Vind Spee
9597	17	18.9389	18.9389	0.84	7.7924
26204	19	24.1222	24.1222	0.46	7.8568
6282	17	20.0889	20.0889	0.6	10.948
92330	18	14.4667	14.4667	0.99	15.4077
51067	17	9.71667	7.40556	0.67	16.4703
91467	17	14.95	14.95	0.72	14.8442
14757	6	19.1889	19.1889	0.62	6.6171
38504	12	-2.17778	-8.51667	0.95	23.6992
85578	6	9.69444	8.95556	0.8	6.7137
70419	19	6.48889	6.48889	0.79	3.0429
47682	19	30.05	30.7556	0.48	22.0248
10886	19	7.22222	4.22778	0.6	17.0821
89795	17	5.11111	2.07778	0.81	13.8943
27528	17	23.3333	23.3333	0.62	6.44

Format Resize Background color Column min/max Save and Close Close

Figure 3: X test

Figure 3 is showing the x test of the given dataset. The given dataset is divided into the 20-30% part into the x test and 70-80% into x train dataset.



y_test - Series

Index	Precip Type
9597	1
26204	1
6282	1
92330	1
51067	1
91467	1
14757	1
38504	2
85578	1
70419	1
47682	1
10886	1
89795	1
27528	1

Format Resize Background color Column min/max Save and Close Close

Figure 4: y test

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 12, Issue 10, October 2023)

Figure 4 is showing the y test of the given dataset. The given dataset is divided into the 20-30% part into the y train dataset.

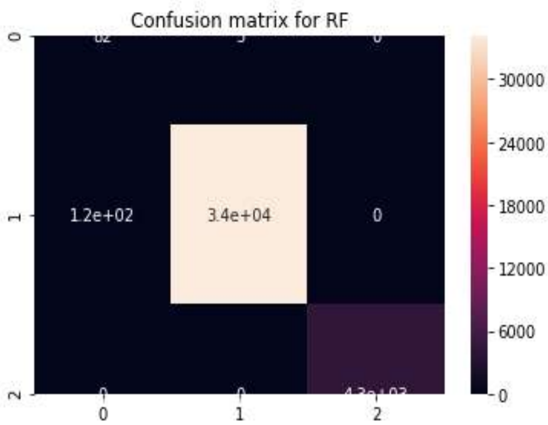


Figure 5: Confusion matrix

Figure 5 presents the confusion matrix of the proposed method. A confusion matrix is a fundamental tool used in machine learning and statistics to evaluate the performance of a classification model.

Table 1:
Result Comparison

Sr. No.	Parameters	Previous Work [1]	Proposed Work
1	Method	LSTM	Random Forest
2	Accuracy (%)	92	99
3	Error Rate (%)	8	1

Table 1 presents a comparison of the previous and proposed work results. The suggested random forest has an accuracy of 99%, whereas the previous LSTM accuracy is 92%. Therefore, it is obvious that the suggested work achieves much better outcomes than prior work.

IV. Conclusion

Weather forecasting is a critical aspect of modern life, with far-reaching implications for agriculture, transportation, disaster management, and various other sectors. This study has explored the application of the Random Forest machine learning technique in the context of weather forecasting, aiming to improve the accuracy and reliability of weather predictions.

The results of this study indicate that Random Forest models excel in capturing complex relationships within meteorological data, leading to accurate forecasts of future weather conditions. By considering the collective knowledge of individual decision trees within the ensemble, Random Forest models offer a powerful solution for tackling the challenges posed by non-linear and dynamic weather patterns.

REFERENCES

[1] Prathyusha, Zakiya, Savya, Tejaswi, N. Alex and S. C C, "A Method for Weather Forecasting Using Machine Learning," 2021 5th Conference on Information and Communication Technology (CICT), Kurnool, India, 2021, pp. 1-6, doi: 10.1109/CICT53865.2020.9672403.

[2] C. Li, M. Zhao, Y. Liu and F. Xu, "Air Temperature Forecasting using Traditional and Deep Learning Algorithms," 2020 7th International Conference on Information Science and Control Engineering (ICISCE), Changsha, China, 2020, pp. 189-194, doi: 10.1109/ICISCE50968.2020.00049.

[3] Y. -L. Li, Z. -A. Zhu, Y. -K. Chang and C. -K. Chiang, "Short-Term Wind Power Forecasting by Advanced Machine Learning Models," 2020 International Symposium on Computer, Consumer and Control (IS3C), Taichung City, Taiwan, 2020, pp. 412-415, doi: 10.1109/IS3C50286.2020.00112.

[4] M. Hostetter and R. A. Angryk, "First Steps Toward Synthetic Sample Generation for Machine Learning Based Flare Forecasting," 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 2020, pp. 4208-4217, doi: 10.1109/BigData50022.2020.9377986.

[5] J. Sleeman, M. Halem, Z. Yang, V. Caicedo, B. Demoz and R. Delgado, "A Deep Machine Learning Approach for Lidar Based Boundary Layer Height Detection," IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium, Waikoloa, HI, USA, 2020, pp. 3676-3679, doi: 10.1109/IGARSS39084.2020.9324191.

[6] A. Doroshenko, V. Shpyg and R. Kushnirenko, "Machine Learning to Improve Numerical Weather Forecasting," 2020 IEEE 2nd International Conference on Advanced Trends in Information Theory (ATIT), Kyiv, Ukraine, 2020, pp. 353-356, doi: 10.1109/ATIT50783.2020.9349325.

[7] A. Golder, J. Jneid, J. Zhao and F. Bouffard, "Machine Learning-Based Demand and PV Power Forecasts," 2019 IEEE Electrical Power and Energy Conference (EPEC), Montreal, QC, Canada, 2019, pp. 1-6, doi: 10.1109/EPEC47565.2019.9074819.

[8] N. L. and M. H.S., "Atmospheric Weather Prediction Using various machine learning Techniques: A Survey," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 422-428, doi: 10.1109/ICCMC.2019.8819643.

[9] C. Feng and J. Zhang, "Reinforcement Learning based Dynamic Model Selection for Short-Term Load Forecasting," 2019 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, 2019, pp. 1-5, doi: 10.1109/ISGT.2019.8791671.

[10] Y. Jiang, J. Yao and Z. Qian, "A Method of Forecasting Thunderstorms and Gale Weather Based on Multisource Convolution Neural Network," in IEEE Access, vol. 7, pp. 107695-107698, 2019, doi: 10.1109/ACCESS.2019.2932027.