RESEARCH ARTICLE

SPECIAL ISSUE

RSP Science Hub

International Research Journal on Advanced Science Hub
2582-4376
Vol. 05, Issue 05S May
www.rspsciencehub.com

Check for updates

http://dx.doi.org/10.47392/irjash.2023.S029

International Conference on intelligent COMPUting TEchnologies and Research (i-COMPUTER) 2023

A Hybrid Approach of Weather Forecasting using Data Mining

Stuti [1], Shashwat Tandon [1], Manjula R [1], Shiv Kumar [1]
[1]School of Computer Science Engineering, Vellore Institute of Technology, Vellore

Email: stutiii811@gmail.com

Article History

Received: 28 February 2023
Accepted: 16 March 2023

Keywords:

Weather forecasting;
Data mining;
Decision Tree;
Gradient Boost;
Gradient descent;
Bagging

Abstract

In the paper, the work focuses on weather prediction by using real time data from day to day. Weather Prediction has proven to be a very important application of Machine Learning since the beginning. Different models were studied and found out ways how prediction could be made more accurate by abandoning the classical models and adopted a hybrid method of including more than hundred decision trees bagged to form an aggregate total. The aggregate results achieved from each tree was considered to be a random split of data, saving a lot of computation time. Gradient Boosting was used to increase accuracy significantly making it a very efficient model to work with. The boosting helped the weak learner Decision Tree to select a random sample of data, fit it with a model and train it sequentially to compensate for the weakness of its predecessor. To improve the accuracy of a model in boosting, a combination of a convex loss function, which measures the gap between the expected and goal outputs, and a penalty term for the complexity of the model were used to reduce a regularized objective function that included both L1 and L2 regression tree functions. The resulting model achieved a significantly high level of accuracy when tested with new data.

1. Introduction

Weather forecasting is a very important aspect of our day to day lives because it enables us to chalk out our activities and make informed decisions. The India Meteorological Department's Yearly Declaration on Climate of India - 2022 states that 2,227 people died in India as a result of severe weather occurrences in 2022, the most in three years (IMD). The technique of predicting the future condition of the atmosphere, including temperature, precipitation, wind, and other meteorological factors, is known as weather forecasting. Although traditional methods of weather forecasting have changed over time, they continue to heavily rely on human discretion and knowledge. The application of data mining models for weather forecasting has attracted a lot of attention recently. Data mining is the process of extracting meaningful information from massive amounts of data using complex algorithms and statistical models. In order to uncover patterns and links within the massive amounts of data, it entails employing computer algorithms to analyse the data. By analysing meteorological data and producing more accurate forecasts, meteorologists can improve forecast accuracy. Data mining techniques have several advantages for weather forecasting. They can handle large datasets, identify complex relationships between weather variables, and can adapt to changing weather conditions. Data mining can also identify patterns that are not immediately visible to the human eye, providing new insights into weather patterns and trends.

In this paper, we analyse the data and forecast the weather using a variety of decision tree techniques. Decision trees are a type of algorithm that break down complex issues into smaller, more manageable portions. (Yonekura, Hattori, and Suzuki) In the context of weather forecasting, a decision tree algorithm would initially consider the present temperature before using this data to anticipate the likelihood of precipitation, such as rain or snow. The algorithm would then consider additional variables like humidity, wind speed, and atmospheric pressure to improve the forecasting accuracy and generate more precise predictions. Decision trees can be considered as flowcharts that make predictions by asking a series of questions. (Barua, Dhoom, and Biswas) says that in weather forecasting, the system makes predictions about the weather by posing inquiries about the current weather and using the responses to do so. Decision trees can be constructed in various ways, such as using information gain or Gini impurity criteria, and can be pruned to prevent overfitting.

(Prasetya and Ridwan) says that one of the advantages of decision tree algorithms is that they can handle missing data and noisy data effectively. Decision trees are also easy to interpret and visualize, which can be useful for understanding the decision-making process and identifying potential areas for improvement. However, there are also some limitations to decision tree algorithms. (Mandale and B.A.) They may not be suitable for complex problems with many variables and may suffer from overfitting if the data is not properly balanced or the tree is not pruned. In addition, (Chauhan and Thakur) says that decision trees may not perform well when the data is highly nonlinear or when there are interactions between variables that are not easily captured by a simple decision tree structure. This paper focuses on exploring how the extremes of rain are related to other factors that cause changes in the weather. Once we identified the most highly correlated factors, we developed a weather prediction model based on them.

This paper presents a weather prediction model that is designed to assist users. We also explore how this model can be enhanced through the use of Gradient Boosting. The paper is structured into several sections. In section 2, we review previous studies related to our work and highlight their limitations. Section 3 outlines the methods we used in this project, while section 4 delves into the technical details of our implementation. The results and implementation are then discussed in section 4. Finally, section 5 presents the results and concludes our work.

2. Existing work and their limitations

Using the linear regression, SMO regression, and M5P model tree data mining algorithms, T. Dananjali, S. Wijesinghe and J. Ekanyake (Dananjali, Wijesinghe, and Ekanyake)forecasted rainfall. After analyzing the data, it was found that the M5P model tree outperformed the other two models in several ways. Specifically, it had the highest directional accuracy, lowest error value, better unpredictability of the error values when compared to the other models, and the strongest correlation between actual and forecasted rainfall quantities. The input and output variables, which correspond to the input parameters and the prediction, are assumed by the linear regression model to have a linear relationship. This assumption prevents it from correctly fitting complex datasets, and as a result, a straight line cannot match the dataset. The observed drawback is that the accuracy of the predictions declines as the forecast period is extended.

The FP Growth Algorithm was utilized by Christy Kunjumon, Sreelekshmi S Nair, Deepa Rajan S, L. PadmaSuresh, and Preetha S L (Sreelekshmi et al.) to classify environmental parameters like maximum and lowest temperatures, rainfall, humidity, etc. As per the survey, various methods and algorithms have been used for weather prediction in the field of data mining. These include supervised and unsupervised machine learning algorithms, artificial neural networks, support vector machines, the FP growth algorithms, K-medoids algorithms, Naive Bayesian classification algorithms, decision tree classification algorithms, and the use of Hadoop with MapReduce. The learning time restriction and the data not fitting into the memory were the main problems their model encountered. It also turned out to be incredibly expensive to build.

Yonekura, K., Hattori, H., and Suzuki, T. (Yonekura, Hattori, and Suzuki) applied a technique for forecasting weather over the next few hours. They are mostly concerned with making predictions within the following hour. Due to a lack of data and computing power, it is typically difficult to predict the weather accurately within a three-hour window; however, there is an increasing demand for very short-term weather forecasts across a variety of industries, including transportation, retail, agriculture, and energy management. The complexity of the weather system, the restricted quantity and quality of data, the difficulty of creating precise short-term forecasts, and the significance of human expertise are the key drawbacks of employing data mining approaches for weather forecasting.

FolorunshoOlaiya, Adesesan, Barnabas Adeyemo (Sagana et al.)applied Artificial Neural Networks (ANN) and the Decision-Tree Algorithms on climatic data. Each algorithm's output was evaluated, and the one that produced the best results was chosen to produce the categorization criteria for the mean weather variables. The absence of data from a longer time span was one of the constraints. A larger data collection, which would include data gathered over many decades, will be required to achieve a better outcome. Among the various recurrent neural network architectures tested, the recurrent TLFD network that utilized the TDNN memory component performed better in training and testing compared to the best TLFD network that used a Gamma memory component. The evaluation of the findings was done using a separate test data set that was created along with the training data, given the limited amount of data available for training and testing. The results were deemed acceptable.

ZhanJie Wang and A. B. M. Mazharul Mujib (Olaiya, Adesesan, and Adeyemo)made use of a model that blended a time series method with a regression approach. This model's primary goal is to forecast the weather for changes in the weather that will have an impact on our everyday lives. This contains the day's high and low temperatures, the likelihood of precipitation, and the speed of the wind. The correlations between meteorological parameters can be found by ANNs, which can then be used to predict the future. ANNs are used to change the topology of the network and to compare how well they perform in training mode. The Dalian Meteorological Bureau has given a time series of meteorological data on which the outcome seen from actual data are based. The test results revealed that the suggested strategy provides an extremely intriguing performance of the implemented network and exhibits very great performance in terms of Mean Squared Error.

Sayantanu Barua, TanniDhoom, Munmun Biswas (Wang and Mujib)use classifiers to predict the weather. Chi square and Naive Bayes algorithms are used for categorization. a sizable dataset was given into the model for training. The user enters data depending on the parameters provided, which is then saved and compared with the dataset included in the model. For identifying user- provided data and predicting the weather, data mining techniques and decision trees are used. A lot of information is required to get better results. The chi-square technique requires a high sample size for an accurate forecast, and if there are more than 20 categories, the model cannot make an accurate prediction. As a result, just a few parameters can be used. The Naive Bayes model will assign a chance of occurrence of zero to any parameter value that the user inputs that was absent during model training, making it unable to predict anything in this regard.

3. PROPOSED WORK

3.1. OVERVIEW

The dataset is divided into a training set and validation set. The Training data: Testing data has a ratio of 80: 20. We make use of the application of bagging decision trees where multiple algorithms are trained independently and combined afterwards to determine the model's average. This technique is useful as it raises the accuracy of the model and reduces the variance factor. This lets us avoid the overfitting problem.

We make use of a hybrid algorithm which is a gradient boosting algorithm that is optimized for designing efficient and scalable machine learning models. It stands for **Extreme Gradient Boosting,** and it can handle large datasets. It efficiently handles missing value problems and is thus well suited for real-time data collection like Weather forecasting. This algorithm makes use of **Decision Trees, concept of Bagging, Gradient Boosting**.

3.1.1. Decision Tree:

The objective of a decision tree algorithm is to make predictions about the value of a target variable by utilizing various input parameters. The algorithm works by dividing the original data set into smaller subgroups based on a test of attribute values. This process is performed recursively on each subset until no further value is added to the predictions or the subset at a given node contains the same value as the target variable. This recursive partitioning approach allows for the creation of a decision tree that can be used for predictive modelling.

3.1.2. Bagging:

This is a classifier that fits the classifiers on each of the subsets of the original dataset and performs individual prediction for each subset. After that it aggregates all the predictions and gives the final prediction. The individual predictions are combined either by using voting or averaging.

3.1.3. Gradient Boosting:

This technique is based on the classification and regression trees where the model recursively executes the process and corrects its predecessor's errors. Each model is trained by using the previous models' errors which are classified as labels and are better optimized in the next model thus boosted.

3.1.4. Gradient Descent:

Gradient descent is an iterative algorithm used to optimize differentiable functions and find their local minimum. The algorithm operates by moving in a series of steps in the direction opposite to the gradient of the function at the current point, which represents the direction of the steepest descent. During each iteration of the algorithm, the cost function is assessed to determine the accuracy of the updated parameter values, and adjustments to the parameters are made until the function is very close to or exactly zero. This ensures that the outcome will have minimal error. It works by subtracting a step size from the present value of the parameter/ feature. This gives the new value of the parameter which is fed into the tree. The calculation of step size is done by multiplying the derivative to the learning rate. The step value shouldn't be too high because doing so could cause it to miss the minimum point and cause the optimization to fail and this value is obtained by a trial-and-error method based on the dataset used and the processing involved.

3.1.5. Extended Gradient Boosting:

This model is an application of bagging gradient boosting decision tree. It also uses the concept of gradient descent. Each independent feature is allocated with some weights or some values before it is fed into the model. The tree then makes predictions and increases the weights of those features whose predictions are incorrect. The second decision tree is then fed with the new values, and this cycle is run again until the required outcomes are achieved. Each model is developed sequentially, and the training is very fast.

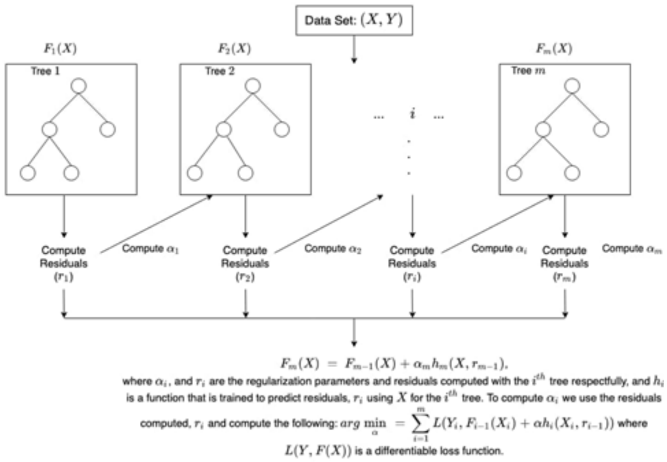

FIGURE 1. Extended Gradient Boosting

Gradient boosting is a machine learning technique for classification that uses regression trees as its weak learners. In this technique, each regression tree maps an input data point to one of its leaves that contains a continuous score. The objective of gradient boosting is to minimize a regularized (L1 and L2) objective function that combines a convex loss function based on the difference between the predicted and target outputs and a penalty term for model complexity. To improve the accuracy of a model in boosting, a combination of a convex loss function, which measures the gap between the expected and goal outputs, and a penalty term for the complexity of the model were used to reduce a regularized objective function that included both L1 and L2 regression tree functions. The resulting model achieved a significantly high level of accuracy when tested with new data. The training process is iterative and proceeds by adding new trees that predict the residuals or errors of prior trees. These new trees are then combined with previous trees to make the final prediction. The term "gradient boosting" comes from the use of a gradient descent algorithm to minimize the loss when adding new models.

L1 Regularization

$$\text{Cost} = \sum_{i=0}^{N}(y_i - \sum_{j=0}^{M} x_{ij}W_j)^2 + \lambda \sum_{j=0}^{M} |W_j|$$

L2 Regularization

$$\text{Cost} = \underbrace{\sum_{i=0}^{N}(y_i - \sum_{j=0}^{M} x_{ij}W_j)^2}_{\text{Loss function}} + \lambda \underbrace{\sum_{j=0}^{M} W_j^2}_{\text{Regularization Term}}$$

FIGURE 2. Formulas of cost for L1 and L2

3.2. Proposed Model

3.2.1. IoT Sensors

Every weather forecasting system makes use of some form of sensors that provide input to the software system for processing and prediction of weather. With constant development of internet of things (IoT) environment monitoring has transformed to Smart Environment Monitoring system where all the parametric data is taken with the use of sensors in web of things. The term ”Internet of Things” (IoT) refers to a collection of tangible objects (referred to as ”things”) that are embedded with sensors, software, and other technologies that allow them to connect and exchange data with other devices and systems over the internet. Multiple sensors are used to get as many data points as possible and get the most accurate data possible. The various sensors involved are as follows:

MQ135 sensor: MQ135 is a type of gas sensor that detects, measures, and monitors a wide range of gasses present in the atmosphere. These include benzene, alcohol, ammonia, carbon dioxide, etc. It is mainly used for testing the air quality, which affects the visibility parameter.

Raindrop Sensor Module: The device in question is capable of measuring the moisture in the air through the use of analog output pins. It generates a digital output when the moisture level surpasses a certain threshold. It is commonly utilized as a rainfall detector and plays a crucial role in agriculture by identifying the moisture level in the soil and regulating the irrigation system accordingly.

BME280 Humidity Sensor: The BME280 device is utilized to determine the temperature, humidity, and pressure of the surrounding environment. The device has an accuracy of $\pm$ 3% for measuring humidity, $\pm$ 1hPa for measuring barometric pressure with absolute accuracy, and a temperature measurement accuracy ranging from $\pm$ 1.0ºC to $\pm$ 1.5ºC.

DHT11 Sensor: The DHT11 sensor is utilized in heating, ventilation, and air conditioning systems to measure humidity and temperature values. It utilizes a capacitive humidity sensor and thermostat to measure the air, and it does not require the use of any analog input pins.

ANDWM3 Anemometer: Anemometer are devices used for measuring the wind speed and wind direction. The most basic is a cup anemometer and others include vane, hot- wire, laser doppler, ultrasonic anemometers all having different implementations for measuring the wind speed and direction.

ESP8266 based Wi-Fi module NodeMCU: This is a Wi-Fi enabled Arduino-link development board that is used for configuring and manipulating all the hardware involved while deploying the sensors.

3.2.2. Consolidating data and preprocessing

Once data has been collected from all the sensors, it has to be processed.This processing involves converting any raw data into comprehensible data. This means that the data is converted to the form that does not contain any noise in the form of blank spaces, gibberish text, errors from the measurement tools, duplicate contradictory values, incomplete data, etc. The data is filtered to contain only numeric values under the correct categories and any non-alpha text, alphabetical data is removed. The outliers are identified and are removed from the dataset.

Encoding of Processed Data: After cleaning the data, One Hot encoding is used to transform the data into a machine-readable format. One Hot encoding method removes the integer encoded variable and adds a new binary variable for each unique integer value. For instance, if the factors are colours such as ”sunny,” ”rainy” and ”windy” they could be encoded as three-component parallel vectors: sunny: [1, 0, 0] (Dananjali, Wijesinghe, and Ekanyake), rainy: [0, 1, 0], windy: [0, 0, 1]. The purpose of encoding data in this way is to convert labels into a numeric format, making it easier for the system to process the large amount of data being fed into it.

Normalization: Once the data is encoded it is then normalized with the help of feature scaling. Here all the numeric data is transformed to fit in a common scale. Our model uses Min-Max normalization where the values of each feature are scaled to a range between 0 and 1 (conventionally). In our case the values are scaled to a range between 0 and 5. Feature scaling lets the model analyse each of the features on the same scale and this helps in significantly boosting the model's performance in terms of computation and processing.

Dropping Null Values: All the null values from each of the features was removed and is visualized in the form of a bar chart. Other data such as duplicate values, missing data were also removed.

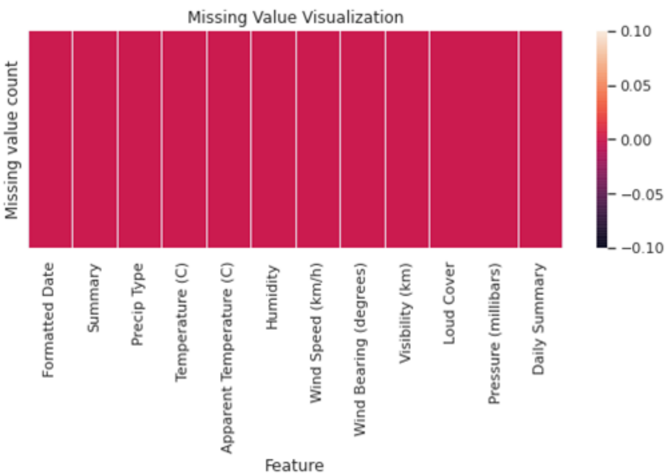

FIGURE 3. **Null Values Visualisation**

3.2.3. Data analysis and feature scaling

Correlation Matrix: A correlation matrix is formed between the Numerical features and the Categorical Features to better visualize the relation between each field. This is then further cleaned by dropping various features and tone down the amount of data to three weather conditions which include, Partly Cloudy, Mostly Cloudy and Overcast.

Features that have a high correlation with the dependent variable are highly dependent on it and have a nearly equal impact on it. We can omit one of the given characteristics when there is a high correlation between the given features.

Outlier Detection and Removal: Once the data is prepared/ cleaned we then move onto pre-processing it. The first step is finding the outliers for each of the

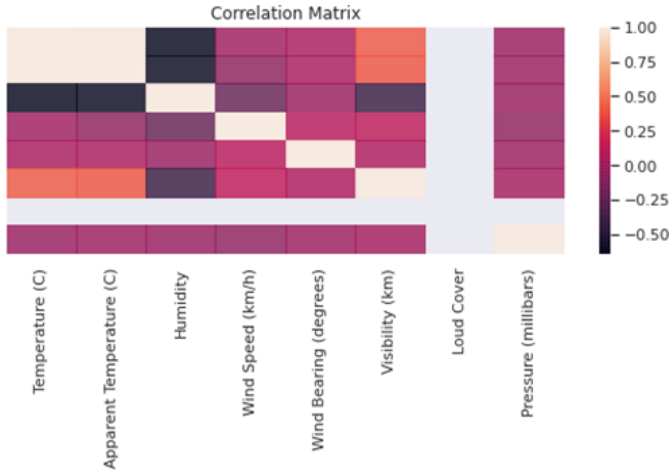

FIGURE 4. **Correlation Matrix**

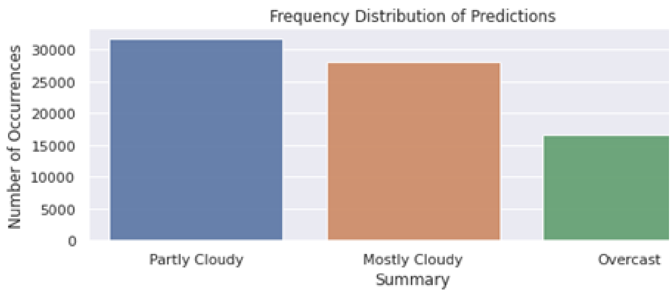

FIGURE 5. **Bar Chart for Frequency distribution**

attributes such as Wind Speed, temperature, humidity, apparent temperature, etc. The outliers are identified using the Quantile Method and are removed.

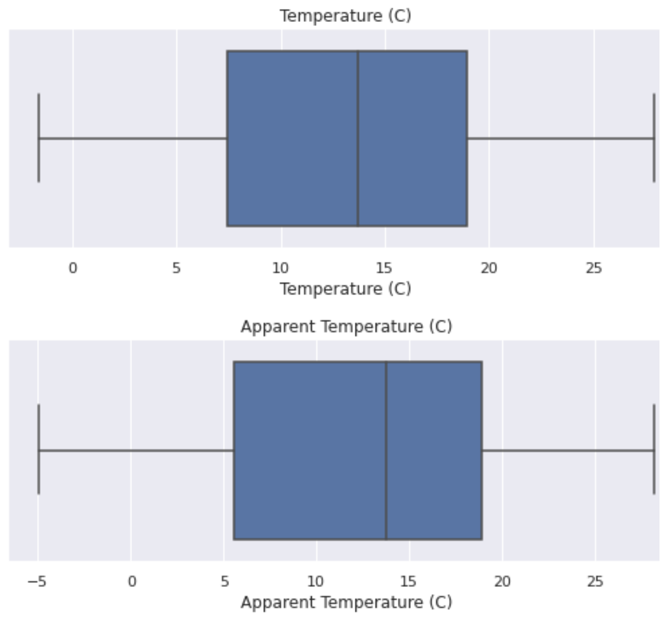

FIGURE 6. **Outlier Detection and Removal**

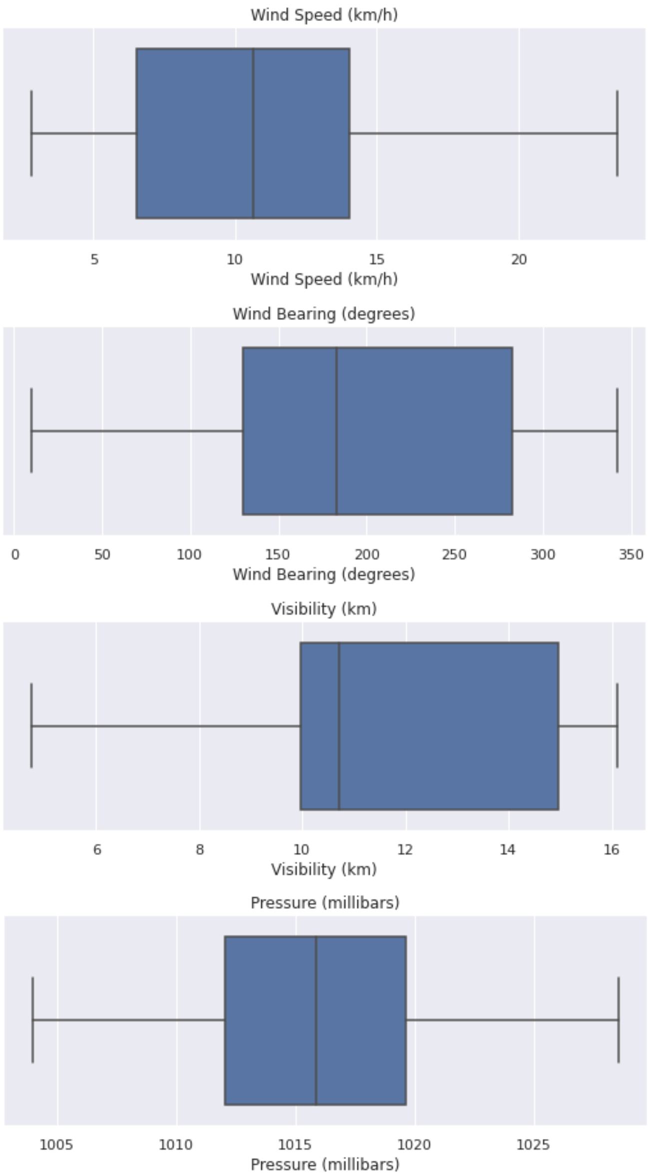

FIGURE 7. Outlier Detection and Removal

FeatureScaling: We scale our data with the help of Min-Max Normalization in the range between [0,5] for each numerical attribute. This step is performed in order to reduce severe deviations caused by some large values of particular parameters. Here the range of independent features is normalized, and this is done for each feature.

3.2.4. Model development

The Make an Initial Prediction and Calculate Residuals: Sum of squares (SS) is a statistical instrument used in regression analysis to measure how well data matches the model and how dispersed the data are. Because it is determined by determining the sum of the squared differences, the sum of squares earned its moniker.

Temperature (C)	Apparent Temperature (C)	Humidity	Wind Speed (km/h)	Wind Bearing (degrees)	Visibility (km)	Loud Cover	Pressure (millibars)
1.876413	1.864407	4.426230	2.737910	3.629518	4.879433	0.0	2.268651
1.856631	1.840074	4.180328	2.773011	3.750000	4.879433	0.0	2.370567
1.860399	2.164793	4.426230	0.269111	2.921687	4.496454	0.0	2.433755
1.675772	1.646249	3.934426	2.734009	3.900602	4.879433	0.0	2.529556

FIGURE 8. Normalized Numerical Values

$$SS_{(residuals)} = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

FIGURE 9. Sum of squares(SS) Formula

The sum of squared residuals can be equal to zero, indicating that the model perfectly fits the data. The accuracy of the model in fitting the data is better when the sum of squared residuals is smaller, and conversely, when the sum of squared residuals is larger, the model is a worse fit for the data.

Build a Gradient Boosted Tree: The model makes predictions or classifications by considering the responses to a previous set of questions. This type of model is a form of supervised learning, meaning it is trained and tested on datasets that already have the desired classifications or outcomes.

Prune the tree: Post-pruning is a technique used to improve the efficiency of decision trees. In this technique, branches from an already fully grown tree are removed to create a pruned tree. This is done to avoid over fitting, which can occur when the decision tree fits the training data too closely and doesn't generalize well to new data. The post-pruning method is demonstrated by the cost complexity pruning algorithm. The pruned node is converted into a leaf node and assigned the class that was most frequent among its parent branches.

Another method to prevent overfitting the data is through pruning. To determine whether a split is legitimate or not, we commence at the bottom of our tree and make our way up. We use gamma to prove the truth (gamma). If Gain is positive, the division is kept; otherwise, it is eliminated.

Calculate the Output values of Leaves:

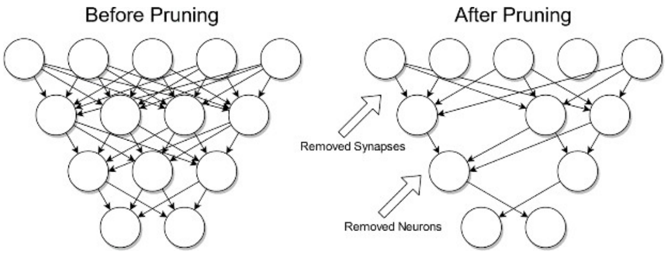

FIGURE 10. **Pruning Mechanism**

$$Output\ Value\ =\ \frac{Sum\ of\ Residuals}{Number\ of\ Residuals\ +\ \lambda}$$

FIGURE 11. **Calculation of output values of leaves**

Now, make new predictions: A machine learning model is treated as a "black box" in a standard optimization process. The model success as determined by the chosen metric is all we are concerned with at each iteration for each chosen collection of hyperparameters. What sort of sorcery takes place inside the dark box is not something we need or want to know. Simply continue to the subsequent iteration, iterate over the subsequent performance assessment, and so forth.

Residual Calculation using new Prediction: We will see that the new residuals are smaller than the old ones, which shows that we've made a modest progress. R residuals will continue to shrink as we repeat this procedure, showing that predicted values are approaching those of the measured values.

We simply keep repeating the same procedure, creating a new tree, making forecasts, and figuring out residuals with each repetition. This is done until the

residuals are extremely tiny, or the algorithm's maximum number of iterations has been achieved.

4. Result analysis and discussion

For Once the Gradient Boost is implemented onto the data, the results are evaluated with the help of Accuracy, Precision, Macro mean, Weighted mean, Recall, F1 Score, Support.

Bagging of trees leads to aggregation of results from various tree algorithms, overcoming the shortcomings of each and balancing it well. The current model initially achieved an accuracy score of 64% but gradient descent boosted the accuracy to a significant 84%.

	precision	recall	f1-score	support
0	0.83	0.83	0.83	3205
1	0.91	0.77	0.83	1485
2	0.83	0.89	0.86	3433
accuracy			0.84	8123
macro avg	0.86	0.83	0.84	8123
weighted avg	0.85	0.84	0.84	8123

FIGURE 12. **Result analysis**

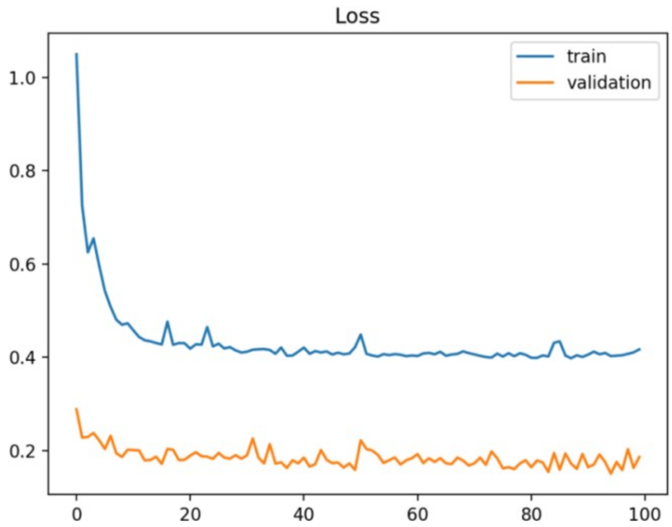

FIGURE 13. **Learning Curve**

This is a Validation Dataset that Is Easier to Forecast Than the Training Dataset of a Train and Validation Learning Curve.

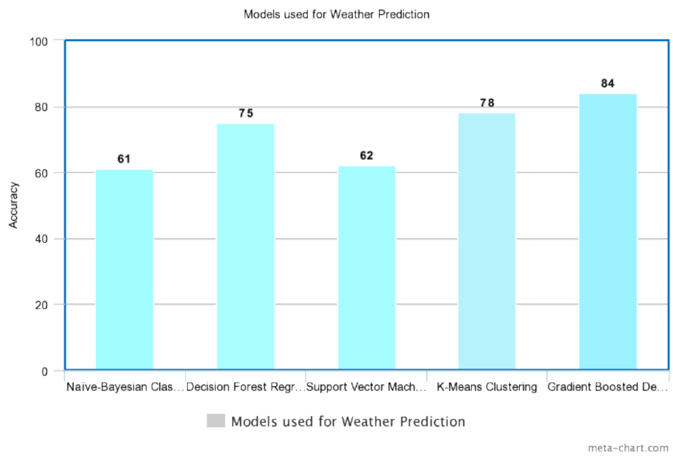

FIGURE 14. **Accuracy plot**

The same dataset was used on different classification models and the Gradient Boosted Variant of Decision Trees performed well among all.

This is an effective Scaling Tree Boosting Technique that has been identified. It has demonstrated exceptional performance in a variety of use cases, including user behaviour analysis, stock sales forecasting, motion recognition, etc. With effective data and memory handling, the system operates significantly quicker on a single machine than any other machine learning method. The algorithm's refining methods increase efficiency and provide speed while using the fewest resources possible.

5. Conclusion

Decision trees thus demonstrate their effectiveness as a tool of decision-making in the prediction of weather. Decision trees are crucial in today's problem-solving tasks like weather forecasting since they are suited for many variable analyses. Bagging of trees leads to aggregation of results from various tree algorithms, overcoming the shortcomings of each and balancing it well. The current model initially achieved an accuracy score of 64% but gradient descent boosted the accuracy to a significant 85%. But this model is limited to structured data. The algorithm is not suitable for handling data that is sparse and unstructured. Furthermore, it is highly vulnerable to outliers, as each classifier is required to correct the errors of the preceding learners in the sequence.

6. Future work

In the Future Work, we can take into consideration the long, short-term memory of data that can increase the prediction probability significantly, with deep learning concepts for higher accuracy. Inclusion of Time Series analysis like ARIMA can help us study the weather trends more accurately. We can make use of efficient Neural Network techniques to identify complex patterns in weather data and make more accurate weather forecasts.

References

Barua, Sayantanu, Tanni Dhoom, and Munmun Biswas. "Weather Forecast Prediction: An Integrated Approach for Analyzing and Measuring Weather Data; SayantanuBarua,TanniDhoom , Munmun Biswas". *International Journal of Computer Applications* (2018): 975–8887.

Chauhan, Divya and Jawahar Thakur. "Data Mining Techniques for Weather Prediction: A Review." *International Journal on Recent and Innovation Trends in Computing and Communication* 2 (2014).

Dananjali, T., S. Wijesinghe, and J. Ekanyake. "Forecasting Weekly Rainfall Using Data mining Technologies". (2020). 10.1109/FITI52050.2020.9424877.

Mandale, Ashwini and Jadhawar B.A. "Weather forecast prediction: a Data Mining application". *International Journal of Engineering Research and General Science* 3.2 (2015).

Olaiya, Folorunsho, Adesesan, and Barnabas Adeyemo. "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies ". *Application of Data Mining Techniques in Weather Prediction and Climate Change Studies By FolorunshoOlaiya,Adesesan Barnabas Adeyemo* (2012). 10.5815/ijieeb.2012.01.07.

Prasetya, Ratih and Anggraeni Ridwan. "Data Mining Application on Weather Prediction Using Classification Tree, Naïve Bayes and K-Nearest Neighbor Algorithm With Model Testing of Supervised Learning Probabilistic Brier Score, Confusion Matrix and ROC RatihPrasetya". *Anggraeni Ridwan JAICT Journal of Applied Communication and Information Technologies* 4.2 (2019).

Sagana, C, et al. "Smart Weather Forecasting using Machine Learning Approach". *2022 6th International Conference on Computing Methodologies and Communication (ICCMC)* (2022): 1072–1078. 10.1109/ICCMC53470.2022.9754024.

Sreelekshmi, S, et al. "Survey on Weather Forecasting Using Data Mining. 2018 Conference on Emerging Devices and Smart Systems". *Survey on Weather Forecasting Using Data Mining. 2018 Conference on Emerging Devices and SmartSystems. Christy Kunjumon* (2018). 10 . 1109 / ICEDSS.2018.8544326.

Wang, Zhanjie and A B M Mazharul Mujib. "The Weather Forecast Using Data Mining Research Based on Cloud Computing." *Journal of Physics: Conference Series* 910 (2017): 012020–012020. 10.1088/1742-6596/910/1/012020.

Yonekura, Kazuo, Hitoshi Hattori, and Taiji Suzuki. "Short-term local weather forecast using dense weather station by deep neural network". *2018 IEEE International Conference on Big Data (Big Data)* (2018). 10.1109/BigData.2018.8622195.

Embargo period: The article has no embargo period.

To cite this Article: , Stuti, Shashwat Tandon , Manjula R , and Shiv Kumar . "A Hybrid Approach of Weather Forecasting using Data Mining." International Research Journal on Advanced Science Hub 05.05S May (2023): 219–228. http://dx.doi.org/10.47392/irjash.2023.S029