Weather Forecasting Using the Random Forest Algorithm Analysis

Abyson Mathew
Department of Computer Application
Amal Jyothi College of Engineering
Kottayam, Kerala, India
abysonmathew@mca.ajce.in

Dr. Juby Mathew
Department of Computer Application
Amal Jyothi College of Engineering
Kottayam, Kerala, India
jubymathew@amaljyothi.ac.in

Abstract- primary sector operations, such as farming, are dependent this weather for productivity. Because the climate is changing at such a rapid rate these days, traditional weather forecast systems are becoming less effective and more time consuming. Improved and dependable weather forecast technologies are necessary to solve these challenges. These forecasts have an impact on a country's economy and people's lives. The major goal of this project is to create a weather forecasting system that can be used in remote places. To forecast weather conditions, data analytics and the machine learning techniques like as random forest classifications is mainly used here. The study recomends a low--cost, portable of weather forecasting systems.

Keywords-Random Forest, weather forecast, machine learning.

I. INTRODUCTION

Many primary sector operations, such as farming, are dependent on the weather for productivity. Because the climate is changing at such a rapid rate these days, traditional weather forecast systems are becoming less effective and more time consuming.

Improved and dependable weather forecast technologies are necessary to solve these challenges. These forecasts have an impact on a country's economy and people's lives. The major goal of this project is to create a weather forecasting system that can be used in remote places. To forecast weather conditions, data analytics and the machines learning techniques like as random forest classifications is mainly used here. The study recomends the low--cost, portable of weather forecasting systems by this.

Random forest classification is a machine learning algorithm that employs an ensemble learning method, in which two or more machine learning models are integrated to generate a single model. It works by creating numerous decision trees while training the dataset and then determining the individual trees' classification modes.

The paper is organized as follows: Section :- II describing the systems that have been implemented till now, while the Section : III was describing the machine learning process in this modules.

II. SYSTEM IMPLEMENTATION

The following steps are commonly used in data mining applications and are included in this paper:

Data Collection and Retrieval- Kaggle provided the data for the research project The CSV file included fields such as Formatted Date, Summary, Temperature in degrees Celsius The apparent temperature in degrees Celsius, Wind Speed is measured in kilometres per hour, Wind Bearing is measured in degrees, and Visibility is measured in kilometres.

WEKA is the most powerful tool for testing machine learning (ML) approaches and applying them to new world data mining problems. It is a collection of data mining-related machine learning methods. The WEKA project's goal is to provide academics with a comprehensive set of machine learning algorithms as well as data pretreatment tools. The algorithms are linked to a database directly. NewWeather2.csv comprises actual data collected throughout a year as well as sample data used for forecast. The data was modelled using the Random-Forest approach. The dataset was then split into two sections: a training set from which the machine can learn and a cross-validation testing dataset The patterns were then documented in order to make more accurate predictions.

Figure:1 WEKA tool front view.

DOI: 10.5281/zenodo.6361990
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

Table 1. Attributes of newWeather.csv

Formatted Date	Date
Summary	String
Temperature (C)	Numeric
Apparent Temperature (C)	Numeric
Humidity	Numeric
Wind Speed (km/h)	Numeric
Wind Bearing (degrees)	Numeric
Visibility (km)	Numeric

Fig : 2

Classifier Output in Random Forest algorithm of Temperature

```
Scheme:       weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:     NewWeather2
Instances:    50
Attributes:   8
              Formatted Date
              Summary
              Temperature (C)
              Apparent Temperature (C)
              Humidity
              Wind Speed (km/h)
              Wind Bearing (degrees)
              Visibility (km)
Test mode:    evaluate on training data

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.02 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds
```

Table 2

Correlation coefficients	0.9965
Mean absolute errors	0.7798
Root mean squared errors	0.9755
Relative absolute error are	20.473 %
Root relative squared error are	22.3539%
Total Number of Instances are	50

Fig : 3

Classifier Output of the Random Forest algorithm in Wind Speed

```
=== Run information ===

Scheme:       weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:     NewWeather2
Instances:    50
Attributes:   8
              Formatted Date
              Summary
              Temperature (C)
              Apparent Temperature (C)
              Humidity
              Wind Speed (km/h)
              Wind Bearing (degrees)
              Visibility (km)
Test mode:    evaluate on training data

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds
```

Table : 3

Correlation coefficients	0.992
Mean absolute errors	1.256
Root mean squared errors	1.58
Relative absolute error are	25.711%
Root relative squared error are	26.35 %
Total Number of Instances are	50

Fig : 4

Classifier Output in Random Forest algorithm of Humidity

```
Scheme:       weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:     NewWeather2
Instances:    50
Attributes:   8
              Formatted Date
              Summary
              Temperature (C)
              Apparent Temperature (C)
              Humidity
              Wind Speed (km/h)
              Wind Bearing (degrees)
              Visibility (km)
Test mode:    evaluate on training data

=== Classifier model (full training set) ===

RandomForest

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.17 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.01 seconds
```

Table 4

Correlation coefficients	0.994
Mean absolute errors	0.033
Root mean squared errors	0.038
Relative absolute error are	23.22 %
Root relative squared error are	23.73 %
Total Number of Instances are	50

III. MACHINE-LEARNING-TECHNIQUE

Random-Forest

The random-forest classification algorithm the machine learning technique that employs ensembling learning to combine two or more machine learning models to produce a new single model.It works by training the dataset with many decision trees and then presenting the categorization modes of the various trees.

Random-Forest believed as the one and only finest ensemble-classifiers in the high-dimensional data. Random-forests are a set of trees predictors in which the values of the randomly sampled vectors with the equall distribution across all trees in the forest is used to build each tree in it. Each trees are trained with replacement on a variety subset of the training datas. To estimate error and variable relevances, the remaining of training data are employed. The

DOI: 10.5281/zenodo.6361990
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam

number of votes from all of the trees is used to assign classes, and the average of the results is utilized for regression.

Algorithm: Each tree is constructed using the following algorithm:

1. Assume the N is the no. of training samples and M is the no. of classifier variables in the data.

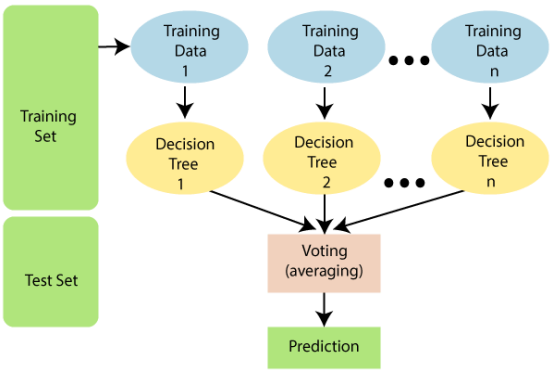
2. We are given the m input variables to use the determining the choices of a tree nodee m should be much smaller than M.

3. To create a training set for this tree, select n times with replacement from the N training examples provided. By predicting the classes of the remaining cases, you can estimate the tree's error.

4. Choose the m variables at random in the each node of trees on to the base of choice at that nodes in it. Calculating the minimum split based on the m variables in the training set.

5. Each trees must be fully mature and hass not been pruned. A fresh sample must be pushed down to the tree for predicting this. It must give the label for the training samples at the terminal node where it ends up. This process must be repeated for each tree in the ensemble, with the end result being that the random forest predictions are based on the average vote of all the trees.

Figure:5



IV. CONCLUSION

Comparative analysis was made by the Random Forest algorithm with the dataset comprising of 8 weather parameters collected over the different period of last year. This outcome is predicted since the Random Forest concept allows it to construct many trees by selecting parameters at random, resulting in a highly accurate classifier. We can forecast the rain for the following year using this random forest technique.

V. REFERENCES

[1] N. Singh, S. Chaturvedi and S. Aakhtaer, "Weather Forecasting Using the Machine-Learning-Algorithm," 2019 International-Conference on the Signal-Processing and Communication in the (ICSC), 2019, 171-174, doi: 10.1109/ICSC45622.2019.8938211.

[2] Weather forecasting using machine learningalgorithm
S. Nagadevi, Varun Ramesh, Helloween James
https://www.psychosocial.com/article/PR260410/14626/

[3] Weather forecast using the sensor data from your IoT hub in Azure Machine Learning.

[4] Smart Weather Forecasting Using Machine Learning: A Case Study in Tennessee
A H M Jakariaa
Smart Weather-Forecasting Using Machine-Learning: A Case-Study in Tennessee

[5] Rohit Kumar Yadav and Ravi Khatri. In 2016. The Weather Forecasting Model using a Data-Mining Technique. International-Journal of the Computer Applications 198,14 7(-2016.)

[6] https;://arnesund.com/2015/05/31/using-amazon-machine-learning-to-predict-the-weather/

DOI: 10.5281/zenodo.6361990
ISBN: 978-93-5607-317-3 @2022 MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam