Weather Prediction Using Random Forest

Bijoy Basak
ID: 2018-2-60-033
Email: 2018-2-60-033@std.ewubd.edu

Simonta Chakraborty
ID: 2018-2-60-036
Email: 2018-2-60-036@std.ewubd.edu

Farhan Ahmed
ID: 2018-2-60-111
Email: 2018-2-60-111@std.ewubd.edu

Abstract: From the last century, one of the most challengeable problem is Weather prediction for both scientifically and technologically all over the world. It is very important in meteorology. For training the model, weather data is used for. Our dataset consists of features like monthly rain and temperature data for a year. We used these data for building our model. We considered rain and temperature data for particular months in a year. We classified our day like normal, dry, wet with normal, cold, warm weather. In our model we used Random forest classifier. We trains the data as per the data collected and predict accuracy, confusion metrics and tree with Random forest classifier. We faced problems for finding accuracy and building confusion metrics in the prediction but we overcame the problems and trying to give the correct prediction with our model.

(I)   INTRODUCTION

In our world, weather forecasting and analysis has become a challenging problem from the last century. There are two main reason: Firstly, it is useful for tourism, natural disaster and most useful for agriculture sector and secondly, it is useful for ongoing improvements in measuring system and the growth of computational power. The weather prediction and the accuracy are the major challenges in our world, which is faced by meteorologist. In the last few years the climate data has been increased though observation records and understudy data and many more sources. This data is more important because its help to find higher accuracy rate to analyse different patterns from the data. Meteorological data is very important and with this data various analysis can be done. It plays an important role to predict temperature, humidity and irrigation system. Here, machine learning algorithms are used to predict weather and collected many historical data. Data from 1901 to 2015 is used in this model. 4 attributes one categorical in weather data class are taken. Data pre-processing is made by handling the null values in the data and handling the outliers. Then we did the explanatory data analysis where we visualize and correlates each attribute and output. Then the graphs are plotted containing all, the attributes and after that we got the important features. Prediction is the last step and prediction is done by machine learning algorithm. We use python as a programming language for climate variation prediction. Factors that contribute to weather variations that affect our daily runtime changes in minimum and maximum temperature, humidity level, rainfall chances

and speed of wind are our primary goal. This prediction model can be used to support many areas like Agriculture, Water Resources, Vegetation and Tourism. Human society is affected in different ways by weather affects. For example, water resources are the main sources of irrigation in production of agriculture crops and amount of rain is an important parameter that affects the crops abruptly due to climate change. It is related to the different human activities. Moreover, due to negative effects of weather resulting in failure of high production, poor growth and low-quality corp. Therefore, changed weather conditions are very risky.

(II)   RELATED WORKS

Weather classification from temperature and rainfall is mostly used to divide climate regions and climate research. It is usually defined in seasonal classifications like: summer, spring, winter or temperature classifications like: dry, humid, warm, cold etc. For both of the classifications temperature and rainfall is not enough. Wind speed, wind direction, humidity etc is necessary especially for the seasonal classifications.

As our dataset only has temperature and rainfall of Bangladesh from the year 1901 to 2015, we have created a few categorical values depending on the temperature range and rainfall range. For temperatures we have three categories: Normal, Cold and Warm. For rainfall we have three categories: Normal, Wet, Dry. To classify the data into the categories we have used the Koppen Climate Regional data.

Bangladesh is in 'Tropical Savanna' climate region. The average temperature is usually from 18°C to 29°C and the average rainfall is from 750mm to 1270mm in a year. We have slightly changed the values to distribute the data among the categories.

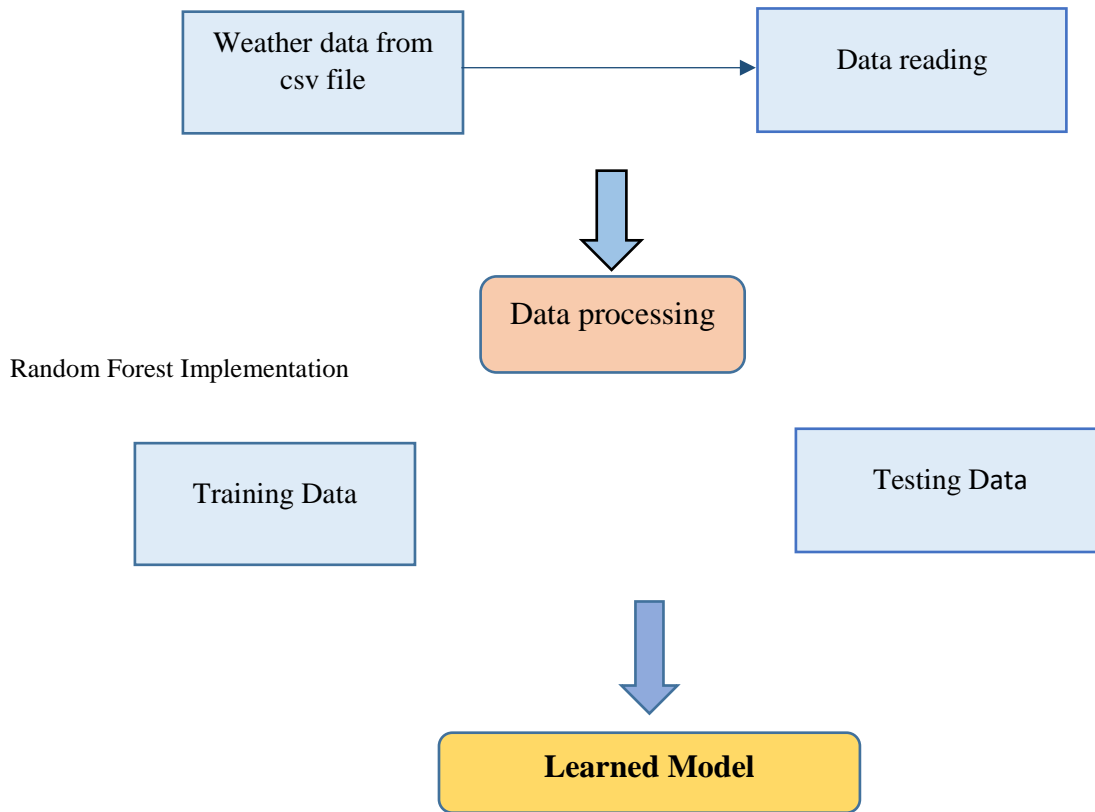# (III)   PROPOSED WEATHER PREDICTION MODEL

Data pre-processing

Fig: Block Diagram

DATASET

We collect our dataset from the website: kaggle in a csv format. We loaded the data in our system for processing. We used the data for rain and temperature from the year 1901 to 2015. Then we create a column weather for creating our categorical values.

DATA NORMALIZATION

In this we divides our dataset into categorical and numerical data. In that weather column we divided like normal, cold and dry weather. We used Wikipedia of Climate for this classifications. Bangladesh's climate is **Tropican Savanna** region. According to **Kopper classification**. **Average Temperature** is around **18°C to 29°C** per year. **Average Rainfall** is around **750 mm to 1270 mm** per year. We slightly changed the parameters and created a few categories.

RANDOM FOREST CLASSIFICATION

Random forest is an ensemble learning algorithm for classification and regression and it is mainly used for classification problems. First, we select the random sample of our dataset then the algorithm constructs a decision tree for every sample, then it gives the prediction result. At last voting will be performed for predicted result and most voted result is the final prediction result among the all predicted results.

TRAINING AND TESTING DATASET

For training the model we had to divide the dataset into training and testing dataset. We did the division in the proportion 80:20. This proportion depends completely on the accuracy that we need. We predict our model from the testing dataset. We trained our model with Random Forest classifier.

ACCURACY AND CONFUSION METRICS

We did the accuracy test and confusion metrics for our dataset using random forest classifier. We got our accuracy. Then we showed our confusion metrics in a plot and finally we showed our prediction with trees.

Confusion metrics:

```
[[1.         0.         0.          0.         0.         0.
   0.         0.         0.        ]
 [0.         1.         0.          0.         0.         0.
   0.         0.         0.        ]
 [0.         0.         1.          0.         0.         0.
   0.         0.         0.        ]
 [0.         0.         0.          0.         0.         0.
   0.         0.         0.        ]
 [0.         0.         0.          0.         0.         0.16666667
   0.         0.         0.        ]
 [0.         0.         0.          0.         0.         0.83333333
   0.         0.         0.        ]
 [0.         0.         0.          0.         0.         0.
   1.         0.         0.        ]
 [0.         0.         0.          0.         0.         0.
   0.         1.         0.        ]
 [0.         0.         0.          0.         0.         0.
   0.         0.         1.        ]]
```

(IV) RESULT

We have used Random Forest Algorithm for the project. Our dataset is small with only 1380 rows of data from 1901 to 2015. Also the categorical values are created by our-self.
In our RF, we have used n-enumerators =80 and max_depth = 4. Any less or more than that reduced the precision.
Precision:

Precision = TruePositives / (TruePositives + FalsePositives)

For our RF, we got the precision of 99.6%. It is very good. As our dataset is relatively small it is quite expected.
From the confusion matrix we see that the Cold Wet weather is predicted as Cold Dry and Warm Normal is predicted as Warm dry. The errors are very low in terms of the 'True Positives'.

Confusion matrix of the classifier
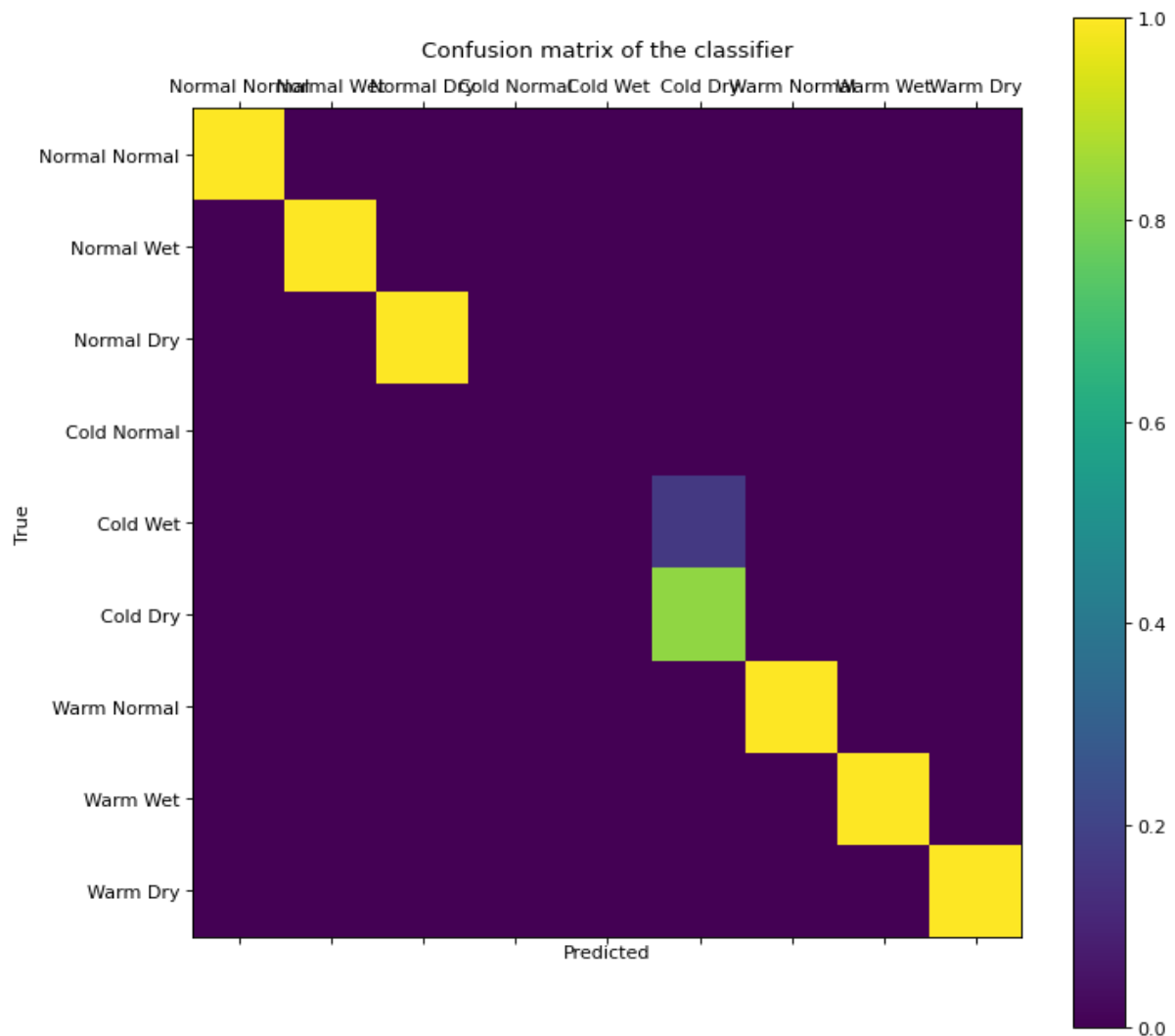
One of the Trees:
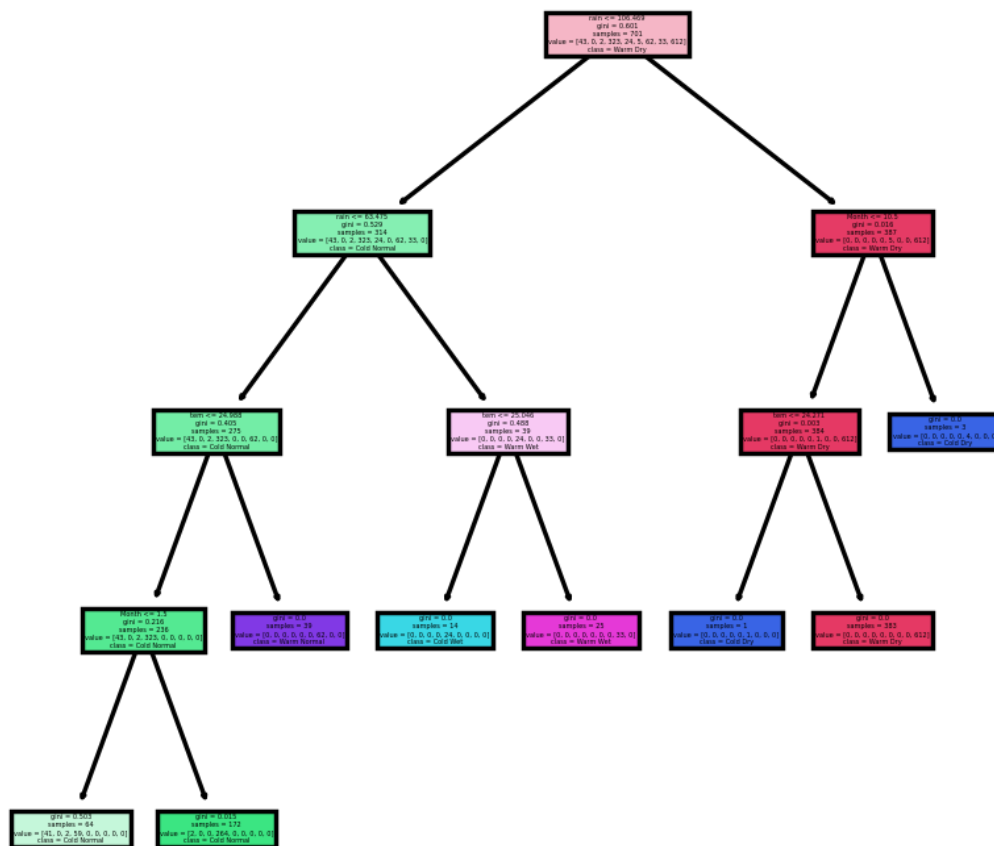
(V)   CONCLUSION

In this project we have used only Random Forest Algorithm to predict the weather using the dataset. The algorithm has done a great work predicting the weather.

References

1. Aishwarya Dhore, Anagha Byakude, Bhagyashri Sonar,Mansi Waste," Weather prediction using the data miningTechniques" IRJET Volume: 04 Issue: 05 | May -2017
2. https://en.wikipedia.org/wiki/Climate
3. https://www.irjet.net/archives/V6/i4/IRJET-V6I41083.pdf