International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 08 | August - 2022 Impact Factor: 7.185 ISSN: 2582-3930

Weather Forecasting using Decision Tree Regression

Amanpreet Kaur[1,*], Meenakshi Sharma[2]

[1]M.tech Research Scholar, CSE Department, RIEIT, Railmajra, SBS Nagar, India

[2]Head of Department, CSE Department, RIEIT, Railmajra SBS Nagar, India

[1]amanpreet16k@gmail.com , [2]13360_meenakshi@rgi.ac.in

ABSTRACT : Weather forecasting [1] is one of the most scientifically and technologically challenging problems around the world in the last century. To make an accurate prediction is indeed, one of the major challenges that meteorologists are facing all over the world. This research work is based on weather prediction using machine learning using the DTR which will help us in getting good accuracy for the weather prediction and prediction of the future weather. The research also demonstrates the existence of a long term trend in the accuracy of the forecasts.

KEYWORDS : LR(Linear Regression), DTR(Decision Tree Regression).

1. INTRODUCTION

1.1 OVERTURE: Weather forecasting entails predicting how the present state of the atmosphere will change. Present weather conditions are obtained by ground observations, observations from ships, observation from aircraft, radio sounds, Doppler radar and satellites. This information is sent to meteorological centers where the data are collected, analyzed and made into a variety of charts, maps and graphs. Modern high-speed computers transfer the many thousands of observations onto surface and upper-air maps. Weather forecasts provide critical information about future weather. There are various techniques involved in weather forecasting, from relatively simple observation of the sky to highly complex computerized mathematical models. Weather prediction could be one day/one week or a few months ahead. The accuracy of weather forecasts however, falls significantly beyond a week. Weather forecasting remains a complex business, due to its chaotic and unpredictable nature. It remains a process that is neither wholly science nor wholly art. The primary aim of the current study is to provide such an assessment to serve:

1. Improvements in weather forecasting using old and huge dataset and showing the various trends.

2. The good accuracy in order to get the perfect mean temperature.

2. BACKGROUND

Stern, H. (2008) reviewed the accuracy of weather forecasts for Melbourne, Australia. He proposed [2] that the analysis shows that skill is evident in forecasts of temperature, rainfall, and qualitative descriptions of expected weather up to 7 days in advance.

Mark Holmstrom, Dylan Liu, Christopher Vo (2016) proposed that Two machine learning algorithms were implemented: linear regression [3] and a variation of functional regression. The input to these algorithms was the weather data of the past two days, which include the maximum temperature, minimum temperature, mean humidity, mean atmospheric pressure, and weather classification for each day. The output was then the maximum and minimum temperatures for each of the next seven days.

Sue Ellen Haupt, Jim Cowie, Seth Linden, Tyler McCandless, Branko Kosovic, Stefano Alessandrini (2018) proposed that the first big advance was in terms of numerical weather prediction (NWP), i.e. integrating the equations of motion forward in time with good initial conditions. But the more recent improvements have come from applying artificial intelligence (AI) techniques to improve forecasting and to enable large quantities of machine-based forecasts.

Tanvi Patil 1, Dr. Kamal Shah2 (2021) proposed LR has been used for forecasting the minimum and maximum temperature and wind speed. The major objectives of Linear Regression: Linear regression has been used for the following two objectives.

In order to find the relationship among variables and to estimate the values of some attributes so that new observations are entertained.

3. PURPOSE

The purpose of the current paper is to show the non-linear trends in order to show the temperature trends since the past years and also to predict the future weather with a good accuracy using Decision Tree Regression algorithm. In the earlier researches where different algorithms have used like LR, Bayesian Networks [4], Neural Networks, Functional Regression [5] etc. The accuracy that we get is quite low than what should we actually expect. We get a very good accuracy with the approach that is being used by us. Decision trees supports non linearity, where LR supports only linear solutions [6]. When there are large number of features with less datasets (with low noise), linear regressions may outperform Decision trees/random forests. For categorical independent variables, decision trees are better than LR. Decision tree builds regression [7] or classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with **decision nodes** and **leaf nodes**. A decision node (e.g., Outlook) has two or more branches (e.g., Sunny, Overcast and Rainy), each representing values for the attribute tested. Leaf node (e.g., Hours Played) represents a decision on the numerical target.

The topmost decision node in a tree which corresponds to the best predictor called **root node**. Decision trees can handle both categorical and numerical data.

4. Results and Discussions

4.1 Data cleansing

Weather data cleaning [8] is fundamental to the provision of high quality weather data. Weather forecasters collect the data for the weather prediction. The dataset which has been used in this model is taken from the government data portal. It contains month wise mean weather all over India. This dataset contains mean temperature of India from the past years.

First of all, the data has been cleaned by applying numerous ways for instance, keeping the data date for January months across all the years, converting string to the date time objects.

4.2 SHOWING THE TRENDS

The trends [9] have shown in every way like

- **a.** Warmest, Coldest, median Monthly Temperature
- **b.** Temperature clusters of Months giving it the interactive different colors for every month of the year
- **c.** Frequency chart of temperature readings
- **d.** Yearly mean temperature
- **e.** Seasonal mean temperature throughout years.
- **f.** Month wise temperature have been shown in an animation frame

The trend can be linear and non-linear. Now, according to our work done above, we come to know that the data is definitely not having he linear trend.

4.3 DECISION TREE REGRESSION:

I am using **Decision Tree Regression [10]** as the data does not actually have a linear trend that we have proved above. This algorithm basically breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. DTR [11] observes features of an object and trains a model in the structure of a tree to predict data in the future to produce meaningful continuous output. Continuous output means that the output/result is not discrete, i.e., it is not represented just by a discrete, known set of numbers or values.

4.3.1 ALGORITHM STEPS:

Fig 4.3.1.1 shows the flowchart of the steps used

a. Importing the libraries: The first step will always consist of importing the libraries that are needed to develop the ML model. The ***NumPy***, ***plotly*** and the ***Pandas libraries*** are imported.

b. Importing the dataset: In this step, we shall use pandas to store the dataset

c. Splitting the dataset into Training Set and Testing Set: In the next step, we have to split the dataset as usual into the *training set* and the *test set*. For this we use test size 0.3 from our dataset which means that this will only be used as test set and the remaining will be used as training set for building the model.

d. Training the decision Tree regression on the Training set: We import the DecisionTreeRegressor class from sklearn.tree and assign it to the variable '***dtr'***. Then we fit the train_x and the train_y to the model by using the dtr.fit function.

After the above steps, we find the accuracy and the accuracy [12] that we get from this model is 96% which is a way more better than all the existing weather prediction models till now. The achievement of this accuracy is because of the huge dataset and the model that we have used for the prediction. The huge the data, the more accurate are the results. With this good accuracy we have predicted the next year data.

ADVANTAGES OF DECISION TREE REGRESSION

a. The decision tree model can be used for both classification and regression problems, and it is easy to interpret, understand, and visualize.
b. The output of a decision tree can also be easily understood.
c. Compared with other algorithms, data preparation during pre-processing in a decision tree requires less effort and does not require normalization of data.
d. The implementation can also be done without scaling the data.
e. A decision tree is one of the quickest ways to identify relationships between variables and the most significant variable.
f. Decision trees are not largely influenced by outliers or missing values, and it can handle both numerical and categorical variables.

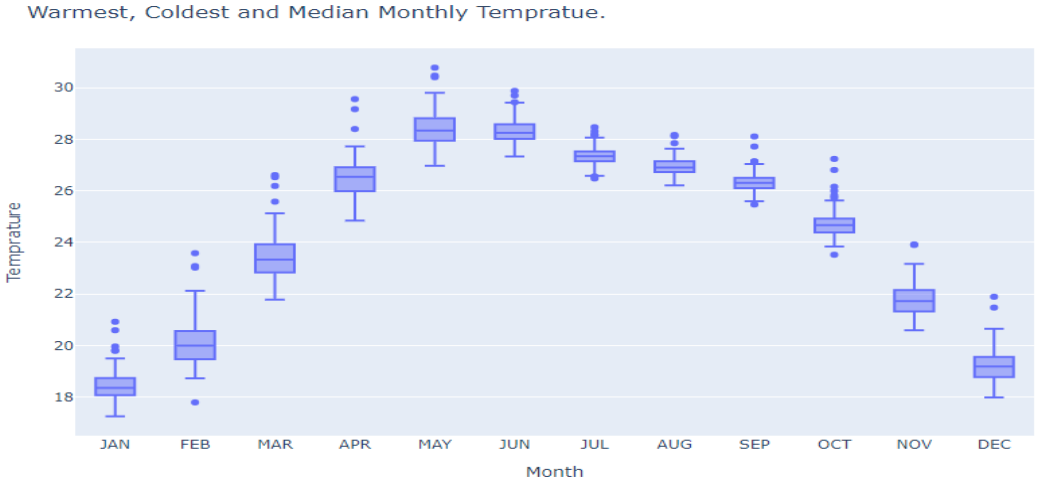
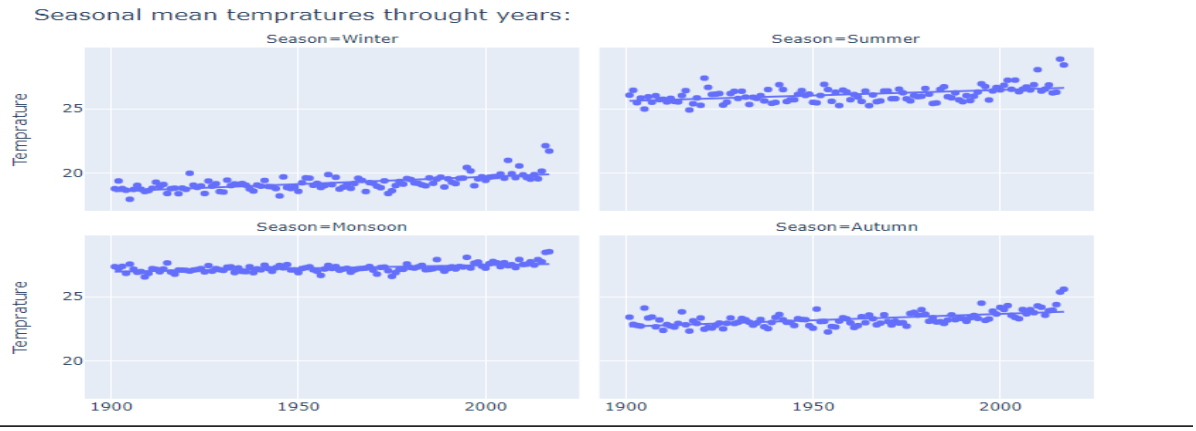


Fig a) shows warmest. Coldest and median monthly temperature which will help us in showing the trends of the temperature throughout the year.



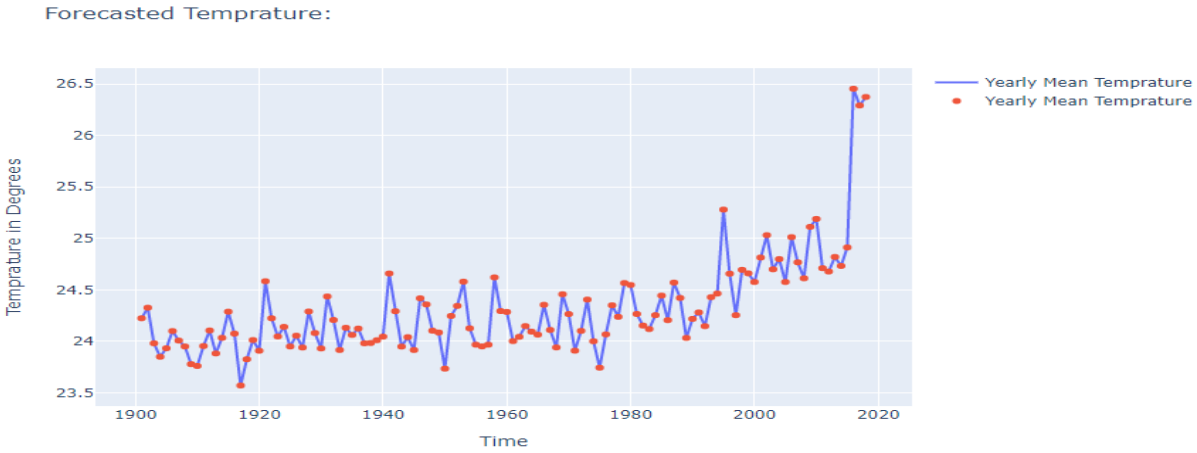


Fig b and c shows seasonal mean temperature throughout the year and Forecasted temperature respectively.
The data that we input is weather dataset from the past years which is to be taken from the govt. website. By inputting the dataset to our model, we can have a mean temperature predicted for the next year.

5. CONCLUSION

This paper documents the non-linear trends in the weather forecasting and also the prediction of the next year with the 96% accuracy using the DTR algorithm of Machine Learning. The mean temperature of the future year is to be predicted. The seasonal weather trends has also been shown which includes summer, winter, monsoon and autumn. Knowing the weather prior will help in many ways in each sector. Weather forecasting is the application of science and technology to predict the state of the atmosphere for a given location. Weather forecasts are made by collecting quantitative data about the current state of the atmosphere and using scientific understanding of atmospheric processes to project how the atmosphere will evolve. There are a variety of end users to weather forecasts. Weather warnings are important forecasts because they are used to protect life and property.

REFERENCES

[1] Mark Holmstrom, Dylan Liu, Christopher Vo, "Machine Learning Applied to Weather Forecasting", Stanford University(Dated: December 15, 2016).

[2] Casta˜n´on, J. (10). Machine Learning Methods that Every Data Scientist Should Know. Consultado em Outubro, 16, 2019

[3] Sue Ellen Haupt, Jim Cowie, Seth Linden" Machine Learning for Applied Weather Prediction" IEEE

[4] Abramson, Bruce, et al." Hailfinder: A Bayesian system for forecasting severe weather." International Journal of Forecasting12.1 (1996): 57-71.

[5] W. Myers, G. Wiener, S. Linden, and S. E. Haupt, "A consensus forecasting approach for improved turbine hub height wind speed predictions,"in Proc. WindPower 2011, Anaheim, CA, May 24, 2011

[6] Tanvi Patil 1, Dr. Kamal Shah2(2021) "Weather Forecasting Analysis using Linear and Logistic Regression Algorithm" Volume: 08 Issue: 06 | June 2021

[7] Stern, H. (2008), "The accuracy of weather forecasts for Melbourne, Australia". Met. Apps, 15: 65?71. doi:10.1002/met.67

[8] Rahm, Erhard and Hong Hai Do. (2000) "Data Cleaning: Problems and Current Approaches." IEEE Bulletin of the Technical Committee on Data Engineering (23): 3-13.

[9] Sushmitha Kothapalli, S. G. Totad, "A Real-Time Weather Forecasting and Analysis", IEEE International Conference on Power, Control, Signals and Instrumentation Engineering (ICPCSI-2017), pp 1567-1570

[10] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. Classification andRegression Trees. CRC Press, 1984

[11] W.-Y. Loh. Regression tree models for designed experiments. Second E.L. Lehmann Symposium, Institute of Mathematical Statistics Lecture Notes-Monograph Series, 49:210-228, 2006

[12] T. Oates and D. Jensen. The effects of training set size on decision tree complexity. In D. H. Fisher, Jr., editor, Proceedings of the Fourteenth International Conference on Machine Learning, pages 254–262, San Francisco, CA, 1997. Morgan Kaufmann.