IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 5, May 2016

Weather Prediction Based on Decision Tree Algorithm Using Data Mining Techniques

Siddharth S. Bhatkande[1], Roopa G. Hubballi[2]

Department of CSE, KLE Dr. MSS College of Engg & Tech. Belgaum India [1]

Professor, Computer Science & Engg, KLE DR M S Sheshgiri College of Engg & Tech., Belgaum[2]

Abstract: "Weather forecasting is a most important application in meteorology and has been one of the most scientifically and technologically challenging problems around the world". We investigate the use of data mining techniques in forecasting attributes like maximum temperature, minimum temperature. This was carried out using Decision Tree algorithms and meteorological data collected between 2012 and 2015 from the different cities. Weather prediction approaches are challenged by complex weather phenomena. Weather phenomena have many parameters like maximum temperature, minimum temperature, humidity and wind speed that are impossible to enumerate and measure. On available datasets we apply the Decision Tree Algorithm for deleting the inappropriate data. Generally maximum temperature and minimum temperature are mainly responsible for the weather prediction. On the percentage of these parameters we predict there is a full cold or full hot or snow fall. This paper develop a model using decision tree to predict weather phenomena like full cold, full hot and snow fall which can be a lifesaving information.

Keywords: Weather Prediction, Data Mining, Decision tree, Meteorological Data Sets.

I. INTRODUCTION

Weather prediction is a challenging task and that too for weather is even more complex, dynamic and mind-boggling. Weather forecast postures right from the antiquated times as a major gigantic undertaking, because it depends on various parameters to predict the dependent variables like maximum temperature, minimum temperature, wind speed and humidity which are changing from weather calculation varies with the some specific location along with its atmospheric attributes. There are many data mining techniques used for weather prediction, but decision tree evaluation can be quantified. Weather forecast is a standout amongst the best natural requirements in our lives. Presently, weather forecasting is made through the application of science and technology. It is made by gathering quantitative information sets about the present condition of the environment through climate station and deciphers by meteorologist.

There are two methods to predict weather
1. Empirical Approach: This approach depends on investigation of past chronicled information of forecast which is gathered in meteorologist's middle and its relationship to an assortment of environmental variables over various parts of areas. The most broadly utilize exact methodologies utilized for climate prediction are Regression, decision tree, artificial neural network, fuzzy logic and group method of data handling [15].

2. Dynamical Approach: This approach, expectations are produced by physical models taking into account arrangement of conditions that anticipate the future climate figure. To foresee the climate by numeric means, meteorologist has create air models that inexact the adjustment in temperature, weight. In our Project climate conjecture expectation is actualized with the utilization of exact measurable method. This paper utilize 4 years (20012-2015) data sets, for example, least temperature, most extreme temperature credits and is going to perform expectation of climate figure utilizing choice tree [15].

II. LITERATURE SURVEY

Literature survey plays a very important role in the project development. Literature survey provides the required knowledge about the project and its background. It also helps in following the best practices in project development. Literature survey also helps in understanding the risk and feasibility of the project. The feasibility of the project depends upon the risk of the project. If the resources, time and money are not available for the project development the risk is higher. Literature survey also gives light on various tools, platforms and operating systems suitable for project development. Once programming begins the programmers require a lot of support and advice.

1. In [5] the author compared the different classification methods namely Decision Trees, Rule-based Methods, Neural Networks, Naïve Bayes, Bayesian Belief Networks, and Support Vector Machines. This paper mentions different notables like decision-tree algorithms like ID3 (Iterative dichotomiser3), C4.5 (successor of ID3), CART (Classification and Regression Tree), CHAID (CHI-squared Automatic Interaction Detector) and MARS (extends decision trees to better handle numerical data.) Here the author concluded that to tap the potential of huge amount of data, decision trees can be used in predicting the dependent variables like fog and rain. Software

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 5, May 2016

equipped with decision tree can provide artificial intelligence to the machine [5].

2. In [6] the author used a decision tree for its simple representation and easy interpretation. Here author says CART is a technique based on collection of rules and values. The author says we can predict the average temperature for the future month if we have the relevant data with a certain degree of accuracy. The model used a decision tree with CART algorithm and was implemented in Weka [6].

3. In [7] the author have used Decision trees to measure the Information gain from various predictors and the split was decided based on nodes having the highest information gain. Over-fitting of the data is prevented by pruning of the tree. By getting rid of nodes that do not contribute much to the information gain, pruning mechanism maximizes information gain. As a result the most effective predictors in a given data set are left behind [7].

4. In this paper used various data mining techniques for prediction of weather forecasting including different classifications like K-Nearest Neighbour, Decision Trees and Naïve Bayes. Decision Trees has achieved quite promising performance among the algorithms. Among the classification Algorithms decision tree has achieved promising results compared to other algorithms. The predicted outcomes are 82.62% of accuracy [8].

5. [9] this paper suggests the algorithm ID3 developed by Ross Quinlan, which is a simple decision tree supervised learning algorithm. To test each attribute at every tree node, the decision tree is constructed by employing a top - down, greedy search algorithm. This paper introduced a new metric named information gain to select the attribute which is useful for classifying the given set [9].

6. Here author concludes that for representations of knowledge discovery, decision trees are fast to execute, very accurate, and best be desired. Classification tasks are adapted by many experts in their own domains and are used by many researchers in various different applications. For constructing more reliable decision trees, multi-tree approaches are the best [10].

7. Here recommends few popular algorithms like CHAID, CART, Quest, and C5.0 for building decision trees [11]. In data mining to analyse the data, to generate the trees and rules, and to make predictions decision tree models are highly recommended.

8. In [6] the author who have proposed a decision tree method to predict weather factors.

9. Many authors conclude that classification and summarization are the two main data mining techniques widely used in weka [12] and Rapid miner tool for weather forecasting.

10. In this paper author describes the capabilities of various algorithms in predicting several weather phenomena such as temperature, windy, humidity, rainfall these parameters concluded that major techniques like decision trees, artificial neural networks, clustering and regression algorithms are suitable to predict weather phenomena. Which shows that decision trees and k-means clustering are best suited data mining techniques for this application [13].

Thus, decision tree is the considered as the powerful solution to the classification problems and it is applied in many real world applications [5]. Many data mining techniques are used for weather forecasting in the present scenario, with various levels of accuracy.
From the Above literature it reveals that there are works which are carried out considering Rule-based Methods, Neural Networks, and Memory based reasoning, Naïve Bayes, Bayesian Belief Networks, and Support Vector Machines. But none of them have attempted identify for Decision tree using data sets hence in this work an attempt is made to predict future weather forecast.

III. WORK FLOW DIAGRAM

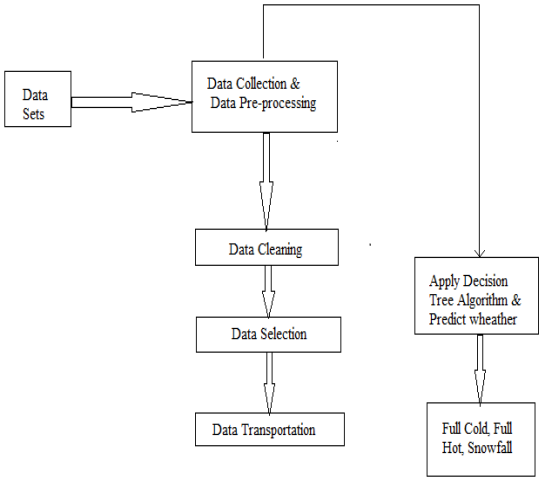

Fig 1. Work Flow Diagram

In this approach paper is completed in four stages as shown in above figure. Data collection and data pre-processing, data cleaning, data selection and finally data transportation. Generally responsible parameters for the weather prediction are maximum temperature and minimum temperature. These are collected from weather department like meteorologists centre and then perform the decision tree algorithm on available datasets and predict the future weather such as day wise or months or years.

A. Data Collection
The data used for this work was collected from meteorologist's centre. The case data covered the period of 2012 to 2015. The following procedures were adopted at this stage of the research: Data Cleaning, Data Selection, Data Transformation and Data Mining.

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 5, May 2016

B. Data Cleaning
In this stage, a consistent format for the data model was developed which is search missing data, finding duplicated data, and weeding out of bad data. Finally system cleaned data were transformed into a format suitable for data mining.

C. Data Selection
At this stage, data relevant to the analysis like decision tree was decided on and retrieved from the dataset. The Meteorological dataset had ten attributes in that were using two attributes for future prediction. Due to the nature of the Cloud Form data where all the values are the same and the high percentage of missing values in the sunshine data both were not used in the analysis.

D. Data Transformation
"This is also known as data consolidation". It is the stage in which the selected data is transformed into forms appropriate for data mining. The data file was saved in Commas Separated Value (CVS) file format and the datasets were normalized to reduce the effect of scaling on the data.

E. Data Mining Stage
The data mining stage was divided into three phases. At each phase all the algorithms were used to analyse the meteorological datasets. The testing method adopted for this research was percentage split that train on a percentage of the dataset, cross validate on it and test on the remaining percentage. There after interesting patterns representing knowledge were identified.

IV. DECISION TREE

A decision tree is a decision bolster device that uses a tree-like graph or model of decisions and their conceivable results, including chance occasion results, asset expenses, and utility. It is one approach to show a calculation. A Decision Tree is a stream outline like tree structure. Each node indicates a test on a property. Every branch speaks to a result of the test. Leaf nodes speak to class appropriation. The choice tree structure gives an express arrangement of "assuming then" guidelines making the outcomes simple to translate [7]. In the tree structures, leaves speak to groupings and branches speak to conjunctions of components that prompt those arrangements. Formally, data increase is characterized by entropy. In other to enhance the precision and speculation of arrangement and relapse trees, different systems were presented like boosting and pruning. Boosting is a procedure for enhancing the precision of a prescient capacity by applying the capacity over and again in an arrangement and consolidating the yield of every capacity with weighting so that the aggregate blunder of the forecast is minimized or growing various free trees in parallel and join them after all the trees have been created. Pruning is completed on the tree to enhance the span of trees and along these lines decrease overfitting which is an issue in extensive, single-tree models where the model starts to fit commotion in the information. At the point when such a model is connected to information that was not used to construct the model, the model won't have the capacity to sum up. Numerous decision tree calculations exist and these include: Alternating Decision Tree, Logit help Alternating Decision Tree (LAD), C4.5 and Classification and Regression Tree (CART).
Decision tree constructs arrangement or relapse models as a tree structure. It separates a datasets into littler and littler subsets while in the meantime a related decision tree is incrementally created. The last 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). Leaf node (e.g., Play) speaks to an order or choice. The highest decision node in a tree which relates to the best indicator called root node. Decision trees can deal with both clear cut and numerical information.

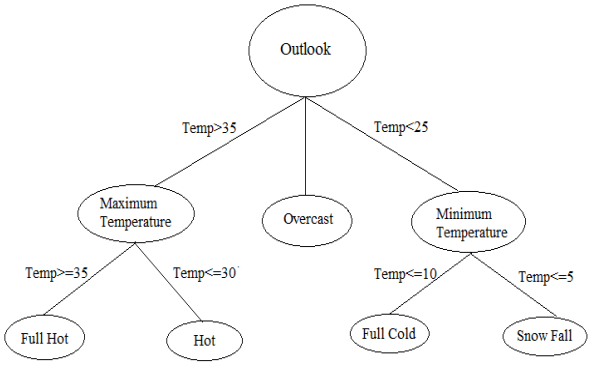

Fig 2. Decision Tree generated by Training data sets

V. DATA SETS

Data Mining is a well known innovation in the field of Data Warehousing and Very Large Data Bases. Data mining is an accumulation of utilized procedures that examine information from a various point of view and speaks to it into helpful data. With a specific end goal to acquired impartial and great data sets, it is gone from different systems. There are numerous associations, for example, Government organizations and some instructive foundations that give access to climate information.

Table 1. Attributes of meteorological Data sets

S.NO	ATTRIBUTE	TYPE	DESCRIPTION
1.	STN	Integer	Station number Of the location
2.	DAY	Integer	Year, month, day
3.	TEMP	Numeric	Mean temperature In F
4.	DEWP	Numeric	Mean dew point In F
5.	SLP	Numeric	Mean sea level pressure In mb
6.	STP	Numeric	Mean station pressure In mb
7.	VISIB	Real	Mean visibility In miles
8.	WDSP	Numeric	Mean wind speed In knots
9.	MXSPD	Numeric	Maximum sustained wind speed In knots
10.	MAX	Numeric	Maximum temperature in F
11.	MIN	Numeric	Minimum temperature In F
12.	PRCP	Binominal	Total precipitation In inches.

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 5, May 2016

NCDC (National Climate Data Centre) and Canadian climatological data centre give a huge climate information range from surface to Radar and Satellite symbolism. Significant information are separated. The chose quality of the meteorological data sets including portrayal of traits alongside their type and brief description are shown in Table 1.

V. WEATHER FORECAST

Weather forecasting is the utilization of science and innovation to anticipate the condition of the environment for a given area. Weather forecasting is a prediction of what the climate will resemble in future days or months or years. Weather forecasting includes a blend of PC models, perceptions, and a learning of patterns and examples. There are an assortment of end uses to climate conjectures. Climate notices are imperative figures since they are utilized to ensure life and property. Forecast in view of temperature and precipitation are vital to horticulture, and along these lines to dealers inside ware markets. Temperature figures are utilized by service organizations to gauge request over coming days. On a regular premise, individuals use climate gauges to figure out what to wear on a given day. Since open air exercises are extremely reduced by overwhelming precipitation, snow and the wind chill, conjectures can be utilized to arrange exercises around these occasions, and to arrange ahead and survive them.

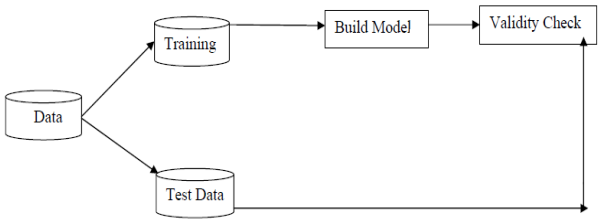

Fig 3. Overview of weather forecast model

VI. PROPOSED WORK

Weather forecasting entails predicting how the future state of the atmosphere will be. There are many ways of obtaining weather conditions values like ground observations, observations from ships and aircraft, Doppler radar, and satellites. “This systems information is sent to meteorological centres where” the data are collected, analysed, and made into a variety of charts, maps, graphs and data sets. Modern high-speed computers transfer the many thousands of observations onto surface and upper-air maps. A final data set is collected for an analysis. Weather forecasting has been one of the” most scientifically and technologically challenging problems around the world in the last century”. “This is due mainly to two factors: first, it’s used for many human activities and secondly, due to the opportunism created by the various technological advances that are directly related to this concrete research field, like the evolution of computation and the improvement in measurement systems. To make an accurate weather prediction is one of the major challenges which is facing meteorologist all over the world. Some people have tried to forecast meteorological characteristics using a number of methods, some of these methods being more accurate than others, this project has been developed for predicting future forecast. By using meteorologist’s data sets a model has been developed based on this attributes like maximum temperature and minimum temperature. The Decision tree algorithm are programmed into a computer and data on the present atmospheric conditions are fed into the computer. The computer solves the algorithm to determine how the different atmospheric variables will change over the next few days or few months or few years.

VII. CONCLUSION

In this paper we used data mining technique and Decision tree algorithm for classifying weather parameters such as maximum temperature, minimum temperature in terms of the day, month and year. The data used from wounder ground weather site between 2012 and 2015 from different cities. The results show how these parameters have influenced the weather observed in these months over the study period. Given enough data the observed trend over time could be studied and important deviations which show changes in climatic patterns identified. Decision trees prove as an effective method of Decision making in Weather prediction. As, decision trees are ideal for multiple variable analyses, it is particularly important in current problem-solving task like weather forecasting. This work is important to climatic change studies because the variation in weather conditions in term of temperature, rainfall and wind speed can be studied using these data mining techniques.

REFERENCES

[1] Application of Data Mining Techniques in Weather Prediction and Climate Change Studies Published Online February 2012 in MECS (http://www.mecs-press.org/) DOI:10.5815/ijieeb.2012.01.07
[2] Ahrens, C. D., 2007, "Meteorology" Microsoft® Student 2008 [DVD], Redmond, WA: Microsoft Corporation, 2007.
[3] Bregman,J.I.,Mackenthun K.M., 2006,Environmental Impact Statements, Chelsea: MI Lewis Publication.
[4] Casas D. M, Gonzalez A.T, Rodrígue J. E. A., Pet J. V., 2009, “Using Data-Mining for Short Term Rainfall Forecasting”, Notes in Computer Science, Volume 5518, 487-490.
[5] Rajesh Kumar, “Decision tree for the weather forecasting”,International Journal of Computer Applications (0975 – 8887) vol.2, August 2013.
[6] Elia Georgiana Petre,“A decision tree for weather prediction”, Seria Matematic - Informatic – Fizic, No.1, pp. 77 – 82, 2009.
[7] Ron Holmes, “Using a decision tree and neural net to identify severe weather radar characteristics”.
[8] Zaheer Ullah Khan and Maqsood Hayat, “Hourly based climate prediction using data mining techniques by comprising entity demean algorithm”, Middle-East Journal of Scientific Research 21 (8): pp. 1295-1300, 2014.
[9] Chandar Sahu, “An intelligent application of fuzzy id3 to forecast seasonal runoff”, International Journal on Cybernetics & Informatics, vol.2, No.1, February 2013.
[10] Quinlan, J. R., “Induction of decision trees. Machine learning”, 1: 81-106, Kluwer Academic Publishers, 1986.
[11] Gurbrinder Kaur, “Meteorological data mining techniques: A survey”, International Journal of Emerging Technology and Advanced Engineering, vol. 2, Issue 8, August 2012.

[12] Mehmed Kantardzic, "Data mining concepts, models, methods and algorithms", IEEE Press 445 Hoes Lane Piscataway, NJ 08854 IEEE press Editorial Board.
[13] Divya Chauhan, Jawahar Thakur, "Data mining techniques for weather prediction: A review", International Journal on Recent and Innovation Trends in Computing and Communication, ISSN: 2321-8169 vol.2 Issue 8.
[14] A.Geetha and Dr.G.M.Nasira, "Implementing service oriented architecture for weather nowcasting", 978-1-4799-3966- 4/14 © 2014 IEEE DOI 0.1109/ICICA 2014.97 pp.444-44.
[15] Jyouti Upadhaya, Assam University ."Climate Change and its impact on Rice productivity in Assam".
[17] http://en.wikipedia.org/wiki/Decision_tree.
[18] http://en.wikipedia.org/wiki/Decision_tree_learning.

BIOGRAPHIES

Siddharth S Bhatkande received B.E. degree in Computer Science & Engg. From Visvesvaraya Technological University of Belgaum in 2013. Currently Pursuing M.Tech degree in Visvesvaraya Technological University of Belgaum. His research interests include Data Mining, Artificial Neural Network and Big Data.

Roopa G. Hubballi received M.Tech in Computer Science & Engg. From Visvesvaraya Technological University of Belgaum. Her area of interest are Image Processing & Data mining

DOI 10.17148/IJARCCE.2016.55114