

# A New Approach for Weather Forecasting in Maharashtra State Using Data-Mining

A.V. Ghumade<sup>1</sup> P.Laddhad<sup>2</sup> P.Kullurkar<sup>3</sup>

<sup>1</sup> Research Scholar <sup>2,3</sup> Assistant Professor

<sup>2,3</sup> Department of Computer Science & Engineering

<sup>1,2,3</sup> Vidharbha Institute of Technology, Nagpur, Maharashtra, India

**Abstract**— In recent past we have observed the problem of Climate change which is badly affecting the agriculture sector of our country and this leads to the decline of economy. Weather forecasting is data mining technique to forecast atmosphere of future time and location related to soil, temperature, heat, cloudiness, wind, crop, dryness, rain etc. In the past many attempts have been made to forecast weather to get better results but it is difficult to get accuracy of results. This paper focuses on upgrading the technique of data mining to next level so that final result should be nearer to real value. There are many techniques of data mining and in this paper data mining algorithms such as Association rule mining & Hybrid model techniques are used weather forecasting to carry agricultural practices in particular region. Thus weather forecasting becomes an important factor for agricultural country like India.

**Key words:** Association Rule Mining (multiple data sets), Hybrid Model, Data Mining

## I. INTRODUCTION

Weather forecasting is one of the best applications of data mining technique. It is the process of collecting data on weather conditions, which includes agriculture, soil, humidity, rainfall, temperature, dew point, wind etc. Weather data can be collected from meteorological data mining, so that this information retrieved and can be transformed into usable knowledge. To get an accurate forecast, a user must have data collected from the past and analysed records. Thus, the weather forecasting is considered as an important application for day-to-day life activities.

Data mining is the process that results in the discovery of new patterns in large data sets. The goal of the data mining process is to extract knowledge from an existing data set and transform it into a human understandable formation for advance use. It is the process of analysing data from different perspectives and summarizing it into useful information. There is no restriction to the type of data that can be analysed by data mining.

Weather forecasting is one of the most challenging factors to occur because there are many problems to get accurate results. In this technique of data mining some classification algorithm like association rule mining and hybrid models are used to detect accuracy problems. Association rule mining means multiple data sets are analysed to get accurate data for weather conditions and agriculture like types of soil, dryness, wind, temperature, cloudiness etc.

Weather information is important for food production, water resource management and all such activity plans that used in day to day life. Weather forecasting is convincing aspect in agronomics states like Maharashtra.

## A. Attributes

Weather is consists of multiple attributes such as air temperature, atmospheric pressure, humidity, precipitation, solar radiation and wind.

It is important to find effective and accurate data to analyse and extract hidden knowledge from the multiple data sets.

## B. Data Mining Concept

It is the concept of abstraction of undetected templates from data storehouses. It is an impressive technology with a great scope to estimate and predict essential information from the databases. Meteorological data are vast, effective, multifarious and high dimensional.

There are different techniques used in data mining as classification, Association Rule Mining, Hybrid Model, Regression, Cluster Analysis, Outlier Analysis and so on. Among these Classification technique is mostly used.

Weather forecast has been one of the most interesting and fascinating domain. The scientists have been trying to forecast the meteorological characteristics using a multiple data set of methods, some of them more accurate than others.

It is the process of analyse the tools to discover patterns and relationships in data that may be used to make valid prediction. Data means collection of information. Database means organized collection of data. Data warehouse means which provides enterprise with memory.

Data Mining- It is derivation of interesting information or patterns from large amount of data. It is an interesting technique that can be implemented in various areas to generate useful information from the existing large volumes of data.

There are two forms of data analysis that can be used for extracting models describing important classes or to predict future data trends. These two forms are as follows –

- Classification
- Prediction

Classification models predict categorical class labels; and prediction models predict continuous valued functions. For example, we can build a classification model to categorize bank loan applications as either safe or risky, or a prediction model to predict the expenditures in dollars of potential customers on computer equipment given their income and occupation.

### 1) Classification

Following are the examples of cases where the data analysis task is Classification

A bank loan officer wants to analyse the data in order to know which customer (loan applicant) are risky or which are safe.

A marketing manager at a company needs to analyse a customer with a given profile, who will buy a new computer.

In both of the above examples, a model or classifier is constructed to predict the categorical labels. These labels are risky or safe for loan application data and yes or no for marketing data.

## 2) Prediction

With the help of the bank loan application that we have discussed above, let us understand the working of classification. The Data Classification process includes two steps –

- 1) Building the Classifier or Model
- 2) Using Classifier for Classification

### a) Building the Classifier or Model:

This step is the learning step or the learning phase.

In this step the classification algorithms build the classifier.

The classifier is built from the training set made up of database tuples and their associated class labels.

Each tuple that constitutes the training set is referred to as a category or class. These tuples can also be referred to as sample, object or data points.

### b) Using Classifier for Classification

In this step, the classifier is used for classification. Here the test data is used to estimate the accuracy of classification rules. The classification rules can be applied to the new data tuples if the accuracy is considered acceptable

## C. Classification and Prediction Methods

Here is the criterion for comparing the methods of Classification and Prediction –

### 1) Accuracy

Accuracy of classifier refers to the ability of classifier. It predicts the class label correctly and the accuracy of the predictor refers to how well a given predictor can guess the value of predicted attribute for a new data.

### 2) Speed

this refers to the computational cost in generating and using the classifier or predictor.

### 3) Robustness

It refers to the ability of classifier or predictor to make correct predictions from given noisy data.

### 4) Scalability

Scalability refers to the ability to construct the classifier or predictor efficiently; given large amount of data.

### 5) Interpretability

It refers to what extent the classifier or predictor understands.

## II. DATA MINING TECHNIQUE

There are few common steps involved in implementing the algorithms as,

- Data Sets
- Data Cleaning
- Attributes Selection
- Association Rule Mining
- Classification Algorithms
- Clustering

1) A **data set** is a collection of data. Most commonly a data set corresponds to the contents of a single database table, or a single statistical matrix, where every column of the table represents a particular variable, and

each row corresponds to a given member of the data set in question. The data set lists values for each of the variables, such as height and weight of an object, for each member of the data set.

- 2) Data cleaning in data mining is a functional step in understanding your data and where its quality issues come from. While data mining is often considered exploratory in nature, data cleaning in data mining is the process of discovering inaccurate or incomplete data—before getting to the analysis and insights. To most cases, data cleaning in data mining is a tedious process that typically requires valuable IT resources, but it's an essential step in evaluating your data. Without proper data quality in regard to the specific initiative that you're executing.
- 3) In data selection there is a huge amount of data available in the Information Industry. This data is of no use until it is converted into useful information. It is necessary to analyse this huge amount of data and extract useful information from it.
- 4) Extraction of information is not the only process we need to perform; data mining also involves other processes such as Data Cleaning, Data Integration, Data Transformation, Data Mining, Pattern Evaluation and Data Presentation. Once all these processes are over, we would be able to use this information in many applications such as Fraud Detection, Market Analysis, Production Control, Science Exploration, etc.
- 5) Association rule mining is primarily focused on finding frequent co-occurring associations among a collection of items. It is sometimes referred to as "Market Basket Analysis", since that was the original application area of association mining. The goal is to find associations of items that occur together more often than you would expect from a random sampling of all possibilities.
- 6) Classification is a data mining function that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. For example, a classification model could be used to identify loan applicants as low, medium, or high credit risks.
- 7) Clustering is the grouping of a particular set of objects based on their characteristics, aggregating them according to their similarities. Regarding to data mining, this methodology partitions the data implementing a specific join algorithm, most suitable for the desired information analysis.

## III. CONCLUSION

Research paper aims at finding suitable data models that achieve a high precision and a high generality with respect to parameters namely rainfall, year, Crop production and area of sowing. For this purpose, different types of Data Mining techniques were evaluated on different data sets. In future work in this paper we plan to capture more details for weather forecasting and agricultural activities. We are moving toward more descriptive traceable accounts that capture not only the basic sources of data and analysers but also more details of how the analysis progressed along with its assumptions and manipulations further work.

REFERENCES

- [1] A survey on Weather forecasting by Data Mining by Audireddy Gayathri<sup>1</sup>, M. Revathi<sup>2</sup>, J. Velmurugan. Sri Venkateswara College of Engineering and Technology, Chittoor and Research Scholar in VIT University, Vellore, India<sup>3</sup> Vol. 5, Issue 2, February 2016.
- [2] Rainfall Prediction using data mining by Sangari.R.S, Dr.M.Balamurugan School of CSE, Bharathidasan University, Trichy, India Vol.2 Issue. 2, February- 2014.
- [3] Alleviating Naive Bayes Attribute Independence Assumption by Attribute Weighting by Nayyar A. Zaidi, Jesus Cerquides, Mark J. Carman Geoffrey I. Journal of Machine Learning Research 14 (2013)
- [4] Data Mining Techniques for Weather Prediction: A Review by Divya Chauhan and Jawahar Thakur. Vol.2 Issue.8, August-2014.
- [5] Classification and Prediction of Future Weather by using Back Propagation Algorithm-An Approach by Sanjay D. Sawaitul, Prof. K. P. Wagh, Dr.P. N. Chatur Government College of Engineering, Amravati, Maharashtra, India. Volumes 2, Issue 1, January 2012).

