International Journal of Research in Advent Technology, Vol.7, No.5S, May 2019
E-ISSN: 2321-9637
Available online at www.ijrat.org

Efficient Analysis of Weather Prediction Using C4.5 Decision Tree and Naive Bayes Algorithm

[1]P.Bamaruckmani, [2]R.Kausalya
Assistant Professor, III-BCA
Department of Computer Applications
Ayya Nadar Janaki Ammal College Sivakasi, Tamil Nadu, India
bamaruckmani@gmail.com

Abstract— Weather prediction is one of the greatest challenges in meteorological department. The main aim of this system is to analyze the performance and efficiency of two algorithms such as C4.5 Decision Tree and Naive Bayes classification. This system can predict the weather condition and generate the rule based itemset by finding information gain using C4.5 decision tree algorithm and probability based using Naïve bayes algorithm. Naive bayes classification technique can used to identify the probabilistic distribution for analyzing temperature and humidity level of specific region. The proposed system is to predict the weather condition of specific region based on the available pre- historical data.

Keywords— Weather Prediction; itemset; information gain; probabilistic distribution.

1. INTRODUCTION

Data mining is a process which finds useful patterns from large amount of data. Data mining can also be defined as the process of extracting implicit, previously unknown and useful information and knowledge from large quantities of noisy, ambiguous, random, incomplete data for practical application.

It is a powerful new technology with great potential to help companies focus on the most important information in their databases. It uses machine learning, statistical and visualization technique to discover and predict knowledge in a form which is understandable to the user. Prediction is the most important technique of data mining which employs a set of pre-classified examples to develop a model that can classify the data and discover relationship between independent and dependent data.

Weather prediction is the application of science and technology to predict the state of the atmosphere for a given location. It is becoming increasingly vital for scientists, agriculturists, farmers, global food security, disaster management and related organizations to understand the natural phenomena to plan and be prepared for the future. The art of weather prediction began with early civilizations using reoccurring astronomical and meteorological events to help them monitor seasonal changes in the weather. Throughout the centuries, attempts have been made to produce forecasts based on weather changes and personal observations.

Weather prediction has been a challenging problem in meteorological department since years. Even after the technological and scientific advancement, the accuracy in prediction of weather has never been sufficient. Even in current date this domain remains as a research topic in which scientists and mathematicians are working to produce a model or an algorithm that will accurately predict weather.

There have been immense improvements in the sensors that are responsible for recording the data from the environment and cancel the noise present in them; along with this new models have been proposed which include different attributes related to weather to make accurate prediction.

Data mining offers a way of analysing data statistically and extract or derive such rules that can be used for predictions. Presently it is being used in many domains such as stock market, sports, banking section, etc. Scientists have now realized that data mining can be used as a tool for weather prediction as well. The basic entity of data mining is data itself.

It is defined as raw set of information which can be used to extract meaningful information depending upon the requirements of the application.

In certain cases regression technique proves to be more effective whereas in other cases, rule based technique and decision tree algorithms give accurate result with a low computational cost. I gave a performance comparison between data mining algorithms like C4.5 decision tree and Naïve Bayes Algorithm includes discussion over the suitability of algorithm when applied to different dataset.

2. REVIEW OF LITERATURE

There are many studies that support the applicability of data mining techniques for weather prediction.

E. G. Petre [1] presented a small application of CART decision tree algorithm for weather prediction. The data collected is registered over Hong Kong. The data is recorded between 2002 and 2005. The data used for creating the dataset includes parameters year, month, average pressure, relative humidity, clouds quantity, precipitation and average temperature. WEKA, open source data mining software, is

used for the implementation of CART decision tree algorithm. The decision tree, results and statistical information about the data are used to generate the decision model for prediction of weather. The way the data is stored about past events is highlighted. The data transformation is required according to the decision tree algorithm in order to be used by WEKA efficiently for weather prediction.

M. A. Kalyankar and S. J. Alaspurkar [2] used data mining techniques to acquire weather data and find the hidden patterns inside the large dataset so as to transfer the retrieved information into usable knowledge for classification and prediction of weather condition. Data mining process is applied to extract knowledge from Gaza city weather dataset. This knowledge can be used to obtain useful predictions and support the decision making process. Dynamic data mining methods are required to build, that can learn dynamically to match the nature of rapidly changeable weather nature and sudden events.

Neha Khandelwal and Ruchi Davey [3] predicted the rainfall of a year by using different 4 climatic factors temperature, humidity, pressure and sea level and thereby using the dataset for calculating drought possibilities in Rajasthan. Certain factors are extracted using data mining techniques. Then correlation analysis is applied on the dataset and correlation is found in the factors. The factors with positive correlations are selected and used for regression analysis. MLR is used for regression analysis for predicting rainfall. Then statistical analysis is applied on that data for finding drought possibility. For drought possibility standard deviation, variance of coefficient, drought indices and drought perception are used. Only one parameter rainfall is considered for analyzing drought condition whereas other climate factors may influence the condition to a wide range. Therefore it is not so accurate.

Abhishek Saxena [4] presented the review of weather prediction using artificial neural networks and studied the benefit of using it. It yields good results and can be considered as an alternative to traditional meteorological approach. The study expressed the capability of artificial neural network in predicting various weather phenomena such as temperature, thunderstorms, rainfall, wind speed and concluded that major architecture like BP, MLP are suitable to predict weather phenomenon. But due to the nonlinear nature of the weather dataset, prediction accuracy obtained by these techniques is still below the satisfactory level.

S. Kannan [5] described empirical method using data mining to make a short term prediction of rainfalls over specific regions. The three months rainfall data of a particular region for five years is analyzed. Accurate and timely weather prediction is a major challenge for research community. Classification technique is used to classify the reason for rainfall in the ground level. Clustering technique is used to group the element that is particular area occupied by rainfall regions and the rainfall is predicted in a particular region. Multiple linear regression models are adopted for prediction but the results give the rainfall data having some approximate value not a predictor value.

Gaurav J. Sawale and Sunil R. [8] Gupta proposed an artificial neural network method for the prediction of weather for future in a given location. Back Propagation Neural Network is used for initial modeling. Then Hopfield Networks are fed with the result outputted by BPN model. The attributes include temperature, humidity and wind speed. Three years data of weather is collected comprising of 15000 instances. The prediction error is very less and learning process is quick. This can be considered as an alternative to the traditional meteorological approaches. Both algorithms are combined effectively. It is able to determine non- linear relationship that exists between the historical data attributes and predicts the weather in future.

P. Hemalatha [6] implemented data mining methods for guiding the path of the ships during sailing. Global Positioning System is used for identifying the area in which the ship is currently navigating. The attributes of weather data includes climate, humidity, temperature, stormy. The weather report of the area traced is compared with the existing database. The analyzed dataset is provided to the decision tree algorithm, C4.5 and ID3. The decision obtained regarding the weather condition is instructed to the ship and the path is hosen accordingly. A close cooperation between the statistical and computational communities provides synergy in data analysis. Few continuous attributes need to be altered as ID3 cannot directly deal with the continuous ranges.

Subana Shanmuganathan and Philip Sallis [7] examined the use of data mining methods to search for the patterns in the adhoc weather conditions, such as time of the day, month of the year, wind direction, speed, and severity using a data set from a single location. The historical weather data, between 2008 and2012 is used from telemetry devices installed in a vineyard in the north of New Zealand. It is shown that using data mining techniques and the local weather condition recorded at irregular intervals can produce new knowledge relating to wind gust patterns for vineyard management decision making. From the data repository, instances relating to the Kumeu River vineyard are extracted for a period of four years (2008–2012). The data collected is cleaned to remove all readings that are outside of Kumeu record readings. The final 86,418 instances and their distribution over the 12 months are presented. The decision tree algorithms used are C5, Quest, CRT and CHAID. SOM is used for the clustering purpose. Multilayered supervised ANN is used for predicting the wind gust. Data mining techniques and statistical methods are run using SPSS. It provides a good tool for analyzing adhoc dataset.

S. S Badhiye [9] used clustering technique with K- Nearest Neighbor method to find the hidden pattern inside the large dataset related to weather so as to transfer the retrieved information into usable knowledge for classification and prediction of climate condition. Temperature and humidity is acquired for a particular time interval. High prediction accuracy is acquired for temperature and humidity. The software can be embedded with the data logger system for the analysis and prediction of parameters in remote areas.

3. METHODOLOGY

Classification of any data is important to know that the data are belongs to anyone class. Based on the results, it is easy to take good decision. But data classification is very complicated process if it is carried out manually. Hence, there is a need for the classification techniques has to be done automatically.

Therefore, machine learning techniques can be used to carry out the data classification operation. The classification is performed with a classifier, which is built from the relevant group of attributes. This classifier is used to classify new data set or test data set based on their attributes. The performance measures such as accuracy, efficiency, etc can be evaluated from the results.

The main objective of this project is to predict the weather status of a particular place using data mining classification algorithm. It has following phases

1. Pre-processing,
2. Feature Selection Methods,
3. Classification.

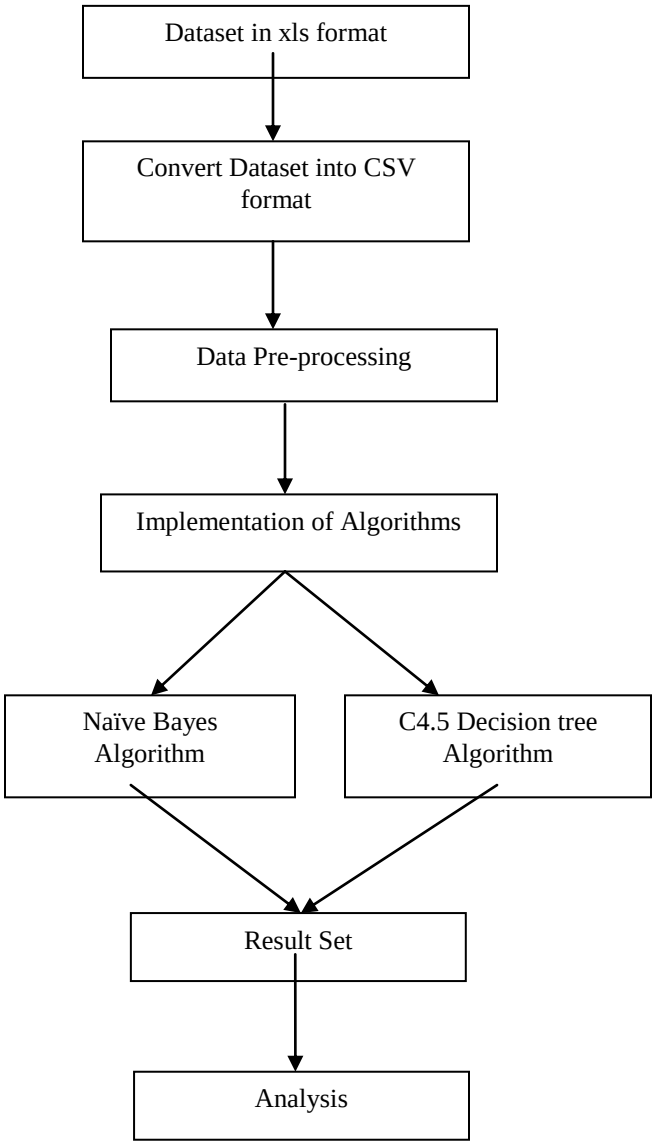

Figure 3.1 Framework of Research Methodology

Dataset Description

This chapter describes about Data Collection and Data Preparation.

Data Collection

The dataset have been collected for Coimbatore city from 1900-2002 from the website "http://www.indiawaterportal.org/met_data/". The dataset have four attributes they are Temperature, Vapour pressure, Cloud cover, Precipitation.

Data Preparation

The weather dataset have been collected and being arranged according to the year and months. Here, all variables are in the form of numerical values. Weather status is the dependent variable and all others are predictor variable.

Table 6.1 Dataset for Weather Prediction

Sno	Year	Month	Temperature	Vapour pressure	Cloud cover	Precipitation	Weather status
1	1901	Jan	Medium	Low	Low	Very low	sunny
2	1901	Feb	Medium	Low	Very low	Very low	sunny
3	1901	mar	High	Medium	Low	Very low	Sunny
4	1901	apr	High	High	Medium	Very low	Sunny
5	1901	may	High	High	Medium	Low	Sunny
6	1901	jun	Medium	High	very high	Medium	Cool
7	1901	jul	Medium	High	very high	Low	Moderate
8	1901	aug	Medium	High	very high	Very low	Moderate
9	1901	sep	Medium	Medium	High	Low	Moderate
10	1901	oct	Medium	Medium	High	Very low	Moderate
11	1901	nov	Medium	Medium	Medium	Low	Sunny
12	1901	dec	Medium	Low	Medium	Very low	Sunny
13	1902	jan	Medium	Low	Low	Very low	Sunny
14	1902	feb	Medium	Low	Very low	Very low	Sunny
15	1902	mar	High	Medium	Low	Very low	Sunny

Pre-processing

Pre-processing can increase the classification accuracy. The proposed system of Pre-processing is to convert numerical data into categorical data based on a certain condition as represented in tables 6.2 to 6.6.

Table 6.2 Attributes Categorical value Description

Attributes	Description	Possible values
Temperature	hotness or coldness of a place	Low, Medium, High
Vapour Pressure	A way of measuring humidity of air	Low, Medium, High
Cloud Cover	Sky obscured by clouds	Very low, Low, Medium, High, Very High
Precipitation	Continual exchange of water between atmosphere and earth surface	Very low, Low, Medium, High, Very High

Table 6.3 Temperature Attribute Category

Possible values	Condition
Low	21 – 24
Medium	24 – 28
High	28 – 32

Table 6.4 Vapour Pressure Attribute Category

Possible values	Condition
Low	21 – 24
Medium	24 – 28
High	28 – 32

Table 6.5 Cloud Cover Attribute Category

Possible values	Condition
Very low	20 – 30
Low	30 – 40
Medium	40 – 60
High	60 – 70
Very high	70 and above

Table 6.6 Precipitation Attribute Category

Possible values	Condition
Very low	0 – 200
Low	200 – 500
Medium	500 – 800
High	800 – 1000
Very high	1000 and above

Feature Selection Methods

Feature Selection Methods are also known as Attribute Selection or variable Selection. It is a process of selecting a subset of relevant features. The main goal of finding a

feature that produces higher classification accuracy. Filter methods are used to select relevant attributes and discard irrelevant attributes. The best method was applied for calculating the accuracy.

There are various feature selection techniques are available in data mining. For this project, Information gain and Gain ratio works well.

Information gain

Information gain is a standard method. The purpose of this technique is to discard irrelevant or relevant features from weather prediction dataset. During information gain feature selection, entropy values have been calculated for whole dataset.

$$\text{Information Gain (D, A)} = \text{Entropy } (D_j) - \sum_{j=1}^{v} \frac{Dj}{D} * \text{Entropy } (D_j)$$

Information can be gained from entropy. Using Information gain feature selection relevant attributes have been selected.

In this research, information gain feature selection is applied to the four attributes are temperature, vapour pressure, cloud cover, precipitation.

Gain ratio

Gain ratio is a modification of the information gain that reduces its bias. It takes the number and size of branches into account when choosing an attribute.

$$\text{Gain Ratio} = \text{Entropy} - \text{Information Gain}$$

It is used to improve the performance of classification algorithm and find out the best feature selection method. Gain ratio is the best feature selection method for weather prediction which selects best attribute.

Classification

The majority of practical machine learning uses supervised learning. Supervised learning is where you have input variables (x) and an output variable (Y) and you use an algorithm to learn the mapping function from the input to the output.

$$Y = f(X)$$

The goal is to approximate the mapping function so well that when you have new input data (x) that you can predict the output variables (Y) for that data.

It is called supervised learning because the process of algorithm learning from the training dataset can be thought of as a teacher supervising the learning process. The algorithm iteratively makes predictions on the training data and this process stops when the algorithm achieves an acceptable level of performance.

Many classification methods are available to predict overall Performance. Classification models include logistic regression, decision tree, random forest, gradient-boosted tree, multilayer perceptron, one-vs-rest, and Naive Bayes.

In that C4.5 is one of the best decision tree classification algorithms. It handles continuous and categorical values. Comparing with ID3, C4.5 has highest speed and pre-pruning. So, the proposed system is carrying out the prediction operation using C4.5 classification.

Attribute Selection is the fundamental step to construct a decision tree. Entropy gain, Information gain, Gain ratio are used to process attribute selection. During, attribute selection C4.5 algorithm selects root node of decision tree.

C4.5 decision tree Algorithm

C4.5 is an algorithm developed by Ross Quinlan that generates Decision Trees, which can be used for classification problems. It improves the ID3 algorithm by dealing with both continuous and discrete attributes, missing values and pruning trees after construction.

Being a supervised learning algorithm, it requires a set of training examples and each example can be seen as a pair: input object and a desired output value. The root of the tree is always the variable which has the minimum value to a cost function. In Decision Trees, it is essential that the nodes are aligned as such that the entropy decreases with splitting downwards.

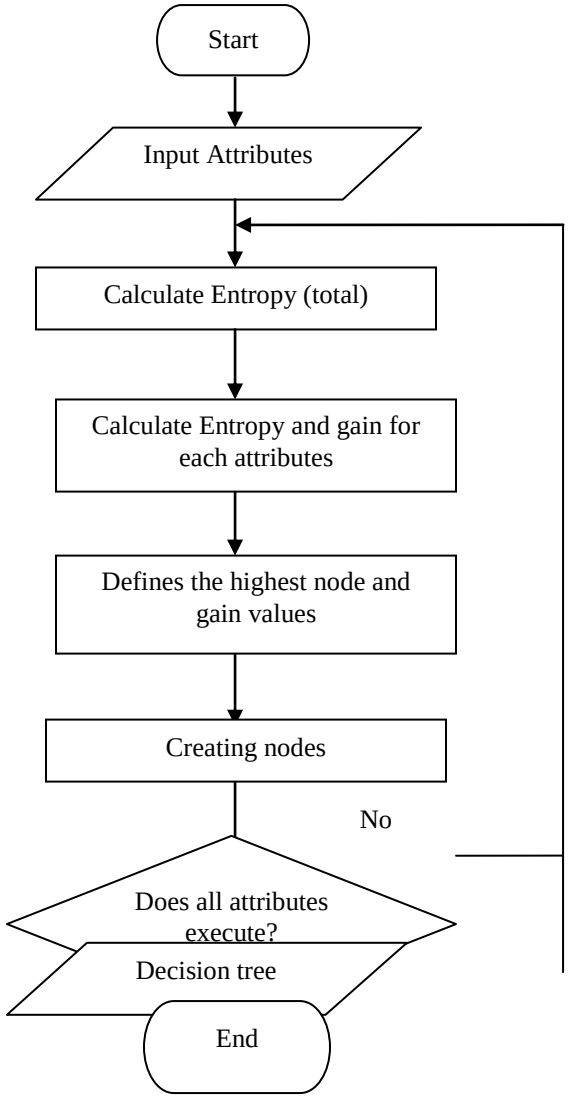

Figure 3.2 Methodology of C4.5 algorithm

A decision tree is a structure that includes a root node, branches, and leaf nodes. Each internal node denotes a test

on an attribute, each branch denotes the outcome of a test, and each leaf node holds a class label. The topmost node in the tree is the root node.

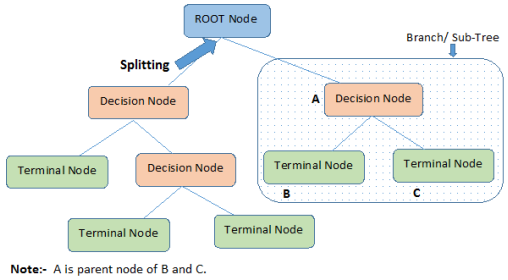

Figure 3.3 decision tree

The algorithm analyzes the training set and builds a classifier that must be able to correctly classify both training and test examples. A test example is an input object and the algorithm must predict an output value.

C4.5 is collection of algorithms for performing classifications in machine learning and data mining. It develops the classification model as a decision tree. C4.5 consists of three groups of algorithm: C4.5, C4.5-no-pruning and C4.5-rules.

J48 is an open source Java implementation of the C4.5 algorithm in the Weka data mining tool.

Improvements from ID.3 algorithm

C4.5 made a number of improvements to ID3. Some of these are

- Handling both continuous and discrete attributes - In order to handle continuous attributes, C4.5 creates a threshold and then splits the list into those whose attribute value is above the threshold and those that are less than or equal to it.
- Handling training data with missing attribute values - C4.5 allows attribute values to be marked as? For missing. Missing attribute values are simply not used in gain and entropy calculations.
- Handling attributes with differing costs.
- Pruning trees after creation - C4.5 goes back through the tree once it's been created and attempts to remove branches that do not help by replacing them with leaf nodes..

Advantages of C4.5 over other Decision Tree systems

- The algorithm inherently employs Single Pass runing Process to mitigate Overfitting.
- It can work with both Discrete and Continuous Data.
- C4.5 can handle the issue of incomplete data very well.

Naive Bayes Algorithm

The Naive Bayes Classifier technique is based on the so-called Bayesian theorem and is particularly suited when the dimensionality of the inputs is high. Despite its simplicity, Naive Bayes can often outperform more sophisticated classification methods.

In machine learning, naive Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong independence assumptions between the features.

It predicts membership probabilities for each class such as the probability that given record or data point belongs to a particular class. The class with the highest probability is considered as the most likely class. This is also known as Maximum A Posteriori.

Naive Bayes is a classification algorithm for binary and multi-class classification problems, is easiest to understand when described using binary or categorical input values.

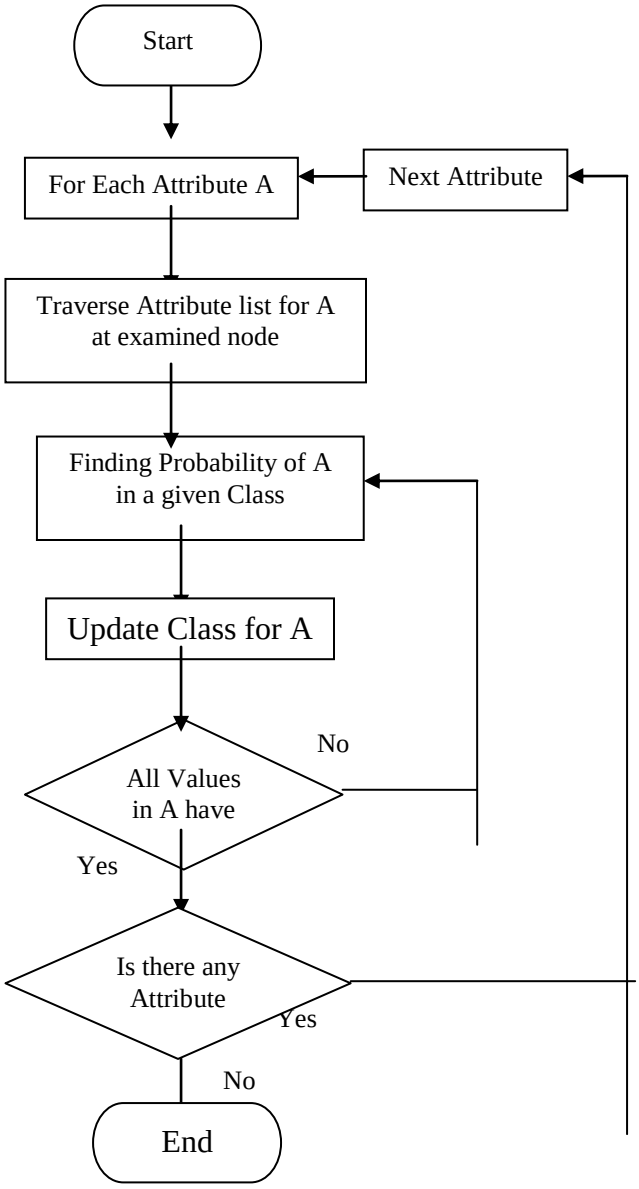

Figure 3.4 Methodology of naïve bayes Algorithm

Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other.

Naive Bayes classifier is a straightforward and powerful algorithm for the classification task. Even working on the dataset with millions of records with some attributes, it is suggested to try Naive Bayes approach.

Principle of Naive Bayes Classifier

A Naive Bayes classifier is a probabilistic machine learning model that's used for classification task. The crux of the classifier is based on the Bayes theorem.

Bayes Theorem

Bayes theorem named after Rev. Thomas Bayes. It works on conditional probability. Conditional probability is the probability that something will happen, given that something else has already occurred. Using the conditional probability, calculate the probability of an event using its prior knowledge.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where,

- P(A) is the probability of hypothesis A being true. This is known as the prior probability.
- P(B) is the probability of the evidence.
- P(B|A) is the probability of the evidence given that hypothesis is true.
- P(A|B) is the probability of the hypothesis given that the evidence is there.

Using Bayes theorem, find the probability of A happening, given that B has occurred. Here, B is the evidence and A is the hypothesis. The assumption made here is that the predictors/features are independent. That is presence of one particular feature does not affect the other. Hence it is called naïve.

In Naïve Bayes Algorithm, the dataset is divided into two parts, namely, feature matrix and the response vector.

- Feature matrix contains all the vectors (rows) of dataset in which each vector consists of the value of dependent features. In the dataset, features are 'Vapour Pressure', 'Temperature', 'Cloud Cover' and 'Precipitation'.
- Response vector contains the value of class variable (prediction or output) for each row of feature matrix. In above dataset, the class variable name is 'Weather Status'.

The fundamental Naive Bayes assumption is that each feature makes an independent and equal contribution to the outcome.

With relation to our dataset, this concept can be understood as

Assume that no pair of features is dependent. For example,

- Temperature being 'High' has nothing to do with the Precipitation or the cloud cover being 'low'. Hence, the features are assumed to be independent.
- Secondly, each feature is given the same weight. For example, knowing only temperature and Precipitation alone can't predict the outcome accurately. None of the attributes is irrelevant and assumed to be contributing equally to the outcome.

Advantages of Naïve bayes Algorithm

- Naive Bayes Algorithm is a fast, highly scalable algorithm.
- Naive Bayes can be use for Binary and Multiclass classification.
- It provides different types of Naive Bayes Algorithms like GaussianNB, MultinomialNB, BernoulliNB.
- It is a simple algorithm that depends on doing a bunch of counts.
- Great choice for Text Classification problems. It's a popular choice for spam email classification.
- It can be easily train on small dataset.

4. RESULTS AND DISCUSSIONS

This chapter describes the experimental results of proposed classification model using data mining techniques. The proposed method is implemented in Weka tool. In that naïve bayes, j48 is used for implementation. This model is used to classify the weather status of a place.

Categorization

Raw data contains numerical data. It needs to be in generalized format. For that numerical data should be transformed into categorical data.

In categorization all the attributes are transformed into categorized format using certain conditions. Categorized data is the preprocessed data. The preprocessed data is taken into the next processes.

Result and Discussions

Accuracy

Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. Accuracy is a great measure but only when symmetric datasets values are of false positive and false negatives are almost same.

Accuracy = sum (diag(mat)) / sum(mat)

Precision

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. High precision relates to the low false positive rate.

Precision = diag(mat) / rowSums(mat)

Recall

Recall is the ratio of correctly predicted positive observations to the all observations in actual class - yes.

Recall = (diag(mat) / colSums(mat)

Result

The analysis of Naïve Bayes and C4.5 Decision Tree algorithm was done simultaneously with dataset containing weather data collected over a period of 102 years. It was found that the performance of C4.5 (J48) decision tree algorithm was far better than that of Naïve Bayes. The accuracy for C4.5 was 97.46% with respect to classifying the instances correctly. On the other hand, Naïve Bayes showed a poor performance of 97.1% while classifying the instances.

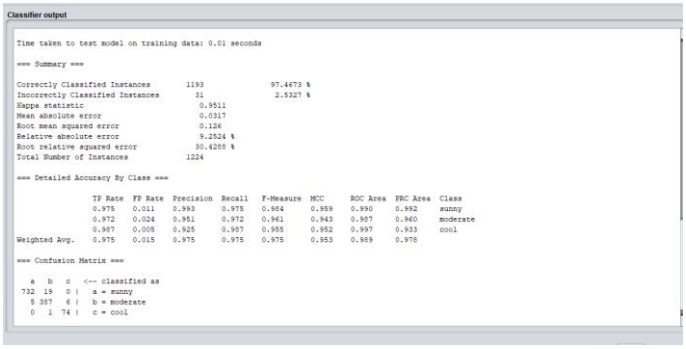

Figure 4.1 Output of C4.5 algorithm

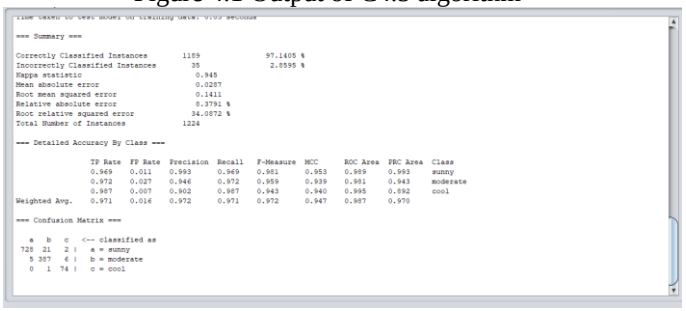

Figure 7.5 Output of C4.5 algorithm

The confusion matrix also supported the above made statement of C4.5 being a better performer in case of weather dataset. The number of instances that were true positives, i.e., true instances and also were predicted true by C4.5 was higher than that of Naïve Bayes and in case of number instances that were true negatives, i.e., false and were predicted as false showed a similar result. Even the precision of C4.5 was considerable higher in this case. Only the time taken to build the model was less in case if Naïve Bayes as compared to C4.5 decision tree.

The performance of C4.5 algorithm improves when the dataset used for application is quite large whereas the performance of Naïve Bayes is comparatively poor. Similarly when the number of attributes increase in the dataset the Naïve Bayes performance drastically affected but C4.5 handles this problem of more number of instances being present in a single dataset in a subtle manner. Therefore it can be said that the performance of C4.5 algorithm was better than that of Naïve Bayes in case of dataset dealing with weather. Further improvements can be made to improve the result of the algorithm by applying appropriate filter to the dataset in pre-processing stage as well as ensemble algorithms can be used along with the C4.5 to achieve a better result.

5. CONCLUSION

The proposed work is intended to find the solution for accurate weather data modeling and prediction using the historical data. Therefore the data mining technique is studied for developing such kind of data model. The data mining techniques analyze the data of some pre-defined pattern and extract the significant information from the data. Weather status has been predicted based on the attributes using C4.5 decision tree Classification and Naïve Bayes Algorithm. The performances of C4.5 decision tree and Naïve Bayes are tested, to predict the weather status. The C4.5 Classification algorithm resulted with 97.38% Classification accuracy and the Naïve Bayes Algorithm resulted with 96.40% Classification accuracy. Comparing C4.5 Decision tree and Naïve Bayes Algorithm, C4.5 Decision tree was the best for my research data.

REFERENCES

[1] Elia Georgiana Petre "A Decision Tree for Weather Prediction", Buletinul, Vol. LXI No. 1, 77-82, 2009.

[2] Meghali A. Kalyankar, S. J. Alaspurkar, "Data Mining Technique to Analyse the Metrological Data", International Journal of Advanced Research in Computer Science and Software Engineering 3(2), 114-118, February – 2013.

[3] Neha Khandelwal, Ruchi Davey, " Climatic Assessment Of Rajasthan's Region For Drought With Concern Of Data Mining Techniques", International Journal Of Engineering Research and Applications (IJERA), Vol. 2, Issue 5, 1695-1697,September-October 2012.

[4] Abhishek Saxena, Neeta Verma, Dr K. C. Tripathi, "A Review Study of Weather Forecasting Using Artificial Neural Network Approach", International Journal of Engineering Research & Technology (IJERT) Vol. 2 Issue 11, November – 2013.

[5] S. Kannan , Subimal Ghosh, "Prediction of daily rainfall state in a river basin using statistical downscaling from GCM output", Springer-Verlag, July- 2010.

[6] P.Hemalatha, "Implementation of Data Mining Techniques for Weather Report Guidance for Ships Using Global Positioning System", International Journal Of Computational Engineering Research Vol. 3 Issue. 3 , march 2013.

[7] Subana Shanmuganathan and Philip Sallis, "Data Mining Methods to Generate Severe Wind Gust Models", 5, 60-80, Atmosphere 2014.

[8] Gaurav J. Sawale, Dr. Sunil R. Gupta, "Use of Artificial Neural Network in Data Mining For Weather Forecasting", International Journal Of Computer Science And Applications Vol. 6, No.2, Apr 2013.

[9] Badhiye S. S., Dr. Chatur P. N., Wakode B. V., "Temperature and Humidity Data Analysis for Future Value Prediction using Clustering Technique: An Approach", International Journal of Emerging Technology and Advanced Engineering, 2250-2459, Volume 2, Issue 1, January 2012.