International Journal of Computer Trends and Technology
ISSN: 2231-2803 / https://doi.org/10.14445/22312803/IJCTT-V67I4P101
Volume 67 Issue 4, 1- 4, April 2019

Original Article

Weather Based Prediction Model for Recommending the Crop Insurance using CART Algorithm

K.P.Mangani[1], Dr.R.Kousalya[2]

[1] *Ph.D. Scholar, PG & Research Department of Computer Science, Research Supervisor*
[2] *Professor & Head, Department of Computer Applications, Dr.N.G.P Arts and Science college,Coimbatore-641048, India.*

Abstract *- In Agriculture, Crop Insurance Prediction is one of the most significant and difficult tasks for researchers in recent years. Weather-based crop Insurance prediction is one of India's most scientifically and technologically challenging problems. This paper investigates the use of data mining techniques in forecasting the insurance model with the attributes like minimum temperature and rainfall with wind speed. This model was carried out using a Decision Tree algorithm like CART with the available meteorological data collected from the Coimbatore district of Tamilnadu. This model is carried out with the CART algorithm and compared with other algorithms, and the performance metrics are recorded.*

Keywords *– Data mining, Classification, Decision Tree, CART, Crop Insurance.*

I. INTRODUCTION

Crop insurance has become an important issue in an agricultural country like India. It plays a vital role in helping our farmer's community. Data mining is discovering and extracting new patterns from large data sets involving functional methods from rainfall Statistics and Meteorological data. Unlike conventional agricultural insurance, weather index insurance does not include any assessment or survey. The mechanism of index insurance is that an indemnity is paid automatically if the observed index becomes under a specific condition. The coverage is based on realizations of a specific weather parameter measured over a pre-specified period at a particular weather station. The insurance can be structured to protect against index realizations that are so high or low that they are expected to cause crop losses [1]. The biggest challenge of weather index insurance is basis risk. It refers to the differences between the actual loss incurred by the farmer and the loss determined by the index entailing claims for non-existent losses and no claims for effective losses [2]. Since the demand for agricultural insurance has increased, a highly accurate weather index with a low basis risk is required. It is concerned with the collection of the information of

Weather conditions and predicts the Banana crop, which is to be insured. Classification and prediction are the techniques used to make out important data classes and predict the feasible solution. This paper formulates the decision tree for the banana crop in the agriculture blocks of Coimbatore districts. It applies the CART algorithm, which has robust learning ability and high speed, and studies the results with accuracy.

II. LITERATURE SURVEY

Karthick and D. Malathi analyze the Naïve Bayes and C4.5 Decision Tree algorithm. It was done simultaneously with the dataset containing weather data collected over 2 years. It was found that the performance of the C4.5 (J48) decision tree algorithm was far better than that of Naïve Bayes to achieve better results [4]

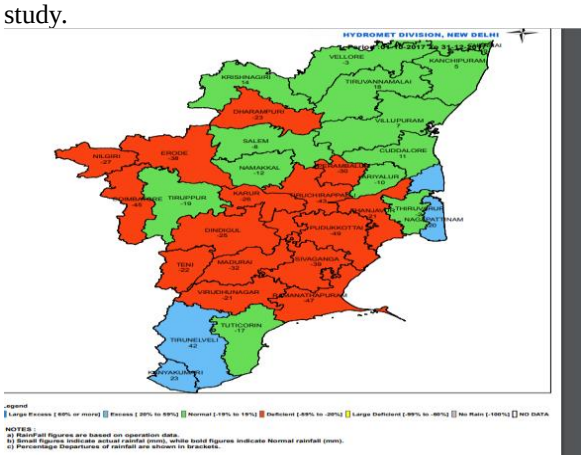
The paper's authors [5] have used a data mining technique and decision tree algorithm for classifying weather parameters such as maximum temperature and minimum temperature in terms of the day, month, and year. The data was used from ground weather sites between 2012 and 2015 from different cities. The results show how these parameters have influenced the weather observed over the study period over these months. Given enough data, the observed trend over time could be studied, and important deviations which show changes in climatic patterns could be identified.

The author explains the huge amount of data given to form a decision tree that can be used in predicting the dependent variable like fog and rain. The software equipped with a decision tree can provide artificial intelligence to the machine. Such software can be used by trekkers, mountaineers, and drivers, facilitating them in decision-making. In this information age, it is quite cost-effective to equip a machine with the sensors and artificial intelligence software so that machine exhibits intelligence [8]

In the paper [6], the author used a decision tree for its simple representation and easy interpretation. The author says CART is a technique based on a collection of rules and values. It also predicts the average temperature for the future month if the workers have the relevant data with a certain accuracy. It also proposes a Computer-aided rule-based rainfall prediction model using CART and C4.5 to find the chance of rain. Then hourly rainfall prediction is performed, rules for rainfall prediction are provided, and then validate the rules. Then ten-fold cross-validation method is performed, and the average accuracies of classifying and predicting raining conditions with CART and C4.5 are 99.2% and 99.3%, respectively. And the average prediction accuracies of estimating hourly rainfall with CART and C4.5 are 92.8% and 93.4%, correspondingly.

The author presents an algorithm for building decision trees in an uncertain environment. This algorithm will use the theory of belief functions to represent the uncertainty about the parameters of the classification problem. This method will be concerned with both the decision tree building task and the classification task [9]

Fig: 1 shows the classification of districts of Tamilnadu based on annual rainfall released by the meteorological department of Chennai, which this paper has taken Coimbatore district of Tamilnadu for study.



Category	Departures from Normals	Colour Code
Large Excess (LE)	60% or more	
Excess (E)	20% to 59%	
Normal (N)	-19% to +19%	
Deficient (D)	-20% to -59%	
Large Deficient (LD)	-60% to -99%	
No Rain	-100%	
No Data	Data Not Available	

Fig. 1 Classification of districts of Tamilnadu based on annual rainfall

III. DATA COLLECTION AND METHODOLOGY

A. Data Collection

The data used for this work was collected from Tamilnadu Agriculture Weather Network, which helps get them with real-time weather data. The case data covered the winter seasonal period of banana of Coimbatore agriculture blocks. The weather dataset had ten (10) attributes. Their type and description are presented in Fig 2, but three parameters are taken for study and applied to the machine learning algorithms used to analyze the datasets. The testing method embraced for this research was a percentage split that trains on a percentage of the dataset, cross-validate on it, and tests on the remaining percentage. Evaluation Metrics were done to select the appropriate algorithms and parameters that make the best model. The following performance metrics were used and compared with other algorithms [3].

Sl.no	Attribute
1	Air Temp(o C) Maximum Minimum
2	Relative Humidity (%)
3	Wind Speed(Kmph)
4	Soil Moisture(%)
5	Soil Temp
6	Rainfall(mm)
7	Solar Radiation (cal/cm2)
8	Atmospheric Pressure (hPa)
9	Leaf Wetness(hr)
10	District

Fig. 2 Weather dataset

B. Machine learning Models

The goal of a machine learning task is to learn the relation between input and output, given a set of training data. For example, given training data (Xi, Yi) where i=1,...,n, and where a pair (Xi, Yi) represents measured environmental parameters and the yield for a certain past season i, the goal is to learn the function f, Y=f (X), which fits best (in a certain sense) the available training data. A possible approach is to find f, which minimizes the average squared error loss but many other forms of losses are also possible. Three machine learning models are widely used and relevant in agricultural applications supervised learning, unsupervised learning, and reinforcement learning; and commonly used machine learning algorithms are Linear Regression, Logistic Regression, Decision Tree, SVM, Naïve-Bayes, KNN, K–Means, and Random Forest Algorithms.

C. Rainfall Prediction Model

The daily weather Report of Coimbatore agricultural blocks was aggregated weekly or monthly. Four months of monsoon (June to September) with an average 650-750 mm usage of seasonal rainfall periods was aimed at making rainfall-based classification to capture adverse events related to rainfall volume and distribution. The weights for different rainfall periods under the weather insurance cover were fixed mainly through statistical optimization techniques. The major objective of these techniques was to either maximize the correlation between rainfall index and area yields or minimize the coefficient of variation (CV) of farmer-level crop revenue per unit area. Farmers and insurance facilitators (mostly NGOs and CBOs) utilize those correlated data for decision-making. The model used here is the decision tree algorithm to predict the insurance based on rainfall and other weather factors.

D. Weather Indices for Banana crop

The insurance agencies must have information about the farmer unit number, crop name with variety, and weather factors with rainfall details. Depending on the type of crop, the importance of attributes changes. For the crop banana, attributes like rainfall, wind speed, and minimum temperature are important. For example, suppose to cultivate bananas; we could target the farmers having fewer water resources and wind speed. Hence the first group of farmers has a constant supply of fresh water for irrigation; the soil with low fertility and an average temperature of 15°C to 35°C is suitable for a banana with relative Humidity of 75-85% and a high velocity of wind which exceeds 80 km /hr. Damages the crop. The decision tree is generated based on the rules framed on temperature, rainfall, and wind speed. And total rainfall details are also needed for framing rules. The typical structure of a weighted rainfall index is of the type.

$$R_t = \sum_{i=1}^{m} w_i r_{it}$$

Where m is the total number of weeks in the growing season, w_i is the weight assigned to the period i of the growing season, and r_{it} the effective rainfall in the period i of year t. The weights w_i are chosen to maximize the sample correlation. The reference or benchmark rainfall index serves as another approach to capturing adverse rainfall to construct an index equal to the maximum consecutive number of *dry days* within a specified period. A dry day is defined as a day with total rainfall below a threshold value. The consecutive Dry day index follows,

CDD Index = Max (No. of Consecutive Days with actual < threshold)

Which is used to frame the rule for the decision tree and the generated pseudo-code for the decision tree.

E. Decision Tree Model

In this model, the prediction of crop insurance is predicted by using minimum, maximum temperature, and seasonal rainfall with the wind speed for 3 months from 2017 – to 2018 for the Coimbatore district of Tamilnadu. Fig: 3 shows that the result is generated in the form of a Decision tree that gives the predicted results as yes or no for crop insurance eligibility.

Fig. 3 Decision tree for banana crop Insurance prediction

IV. RESULTS AND DISCUSSION

Weka classifier is used to study the model and classify the results under different performance measures, as shown in Table 1. R is a programming language and software environment for statistical analysis, graphics representation, and reporting. It is mostly used in Data Mining applications. A decision tree can be created using the 'R' packages by taking the input data in a .csv file and generating the results in the form of a tree[Fig:3]. The nodes in the graph represent an event or choice, and the graph's edges represent the decision rules or conditions. The data is classified less time, and attribute usage is very high. Generally, a model is created with observed data, also called training data. Then a set of validation data is used to verify and improve the model. It gives better accuracy in the prediction of Insurance in Coimbatore blocks. The result is tested, and the predicted results are given below with the same model with different algorithms for the given parameters, shown in Table 1 and Table 2, which are done in weka classifier and R Programming, respectively.

Table 1. Predicted Classification Results in the weka tool

Dataset Used	Weather Data-160 instances			
Measures	Classifiers used in weka			
	Rtree	Decision table	J R I P	J48
Accuracy	95	97.5	93.75	100
Error	0.107	0.104	0.10	0.08
Kappa	0.94	0.937	0.749	0.88
W-M Recall	97.2	95.5	81.1	91.2
W-M Precision	95.3	97.5	93.2	84.4

Table 2. Predicted Classification Results in R tool

Dataset Used	Weather Data-160 instances			
Measures	Classifiers used in R Programming			
	CART	Naive Bayes	SVM	J48
Accuracy	98	97.5	96	97
Error	0.06	0.05	0.09	0.08
Kappa	0.94	0.97	0.86	0.92
W-M Recall	97.2	96.5	87.1	96.2
W-M Precision	98.5	96.5	91.2	93.4

And the diagrammatic view of the predicted results is given in the figure.

The graph shown in Fig 4 shows the performance measures of the classification model in the Weka tool.

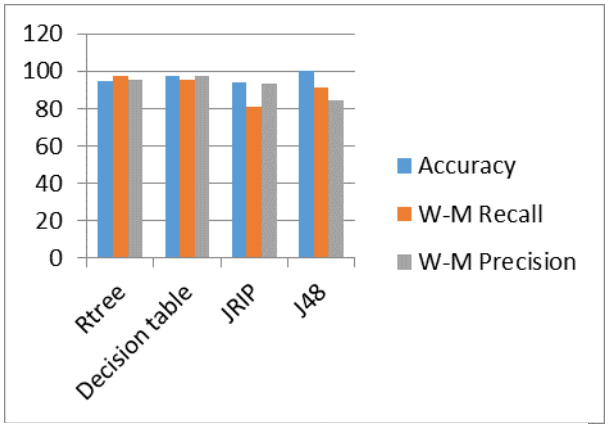


Fig. 4 Weka-Performance Measures

The graph shown in Fig 5 shows the performance measures of the classification model in the R tool.

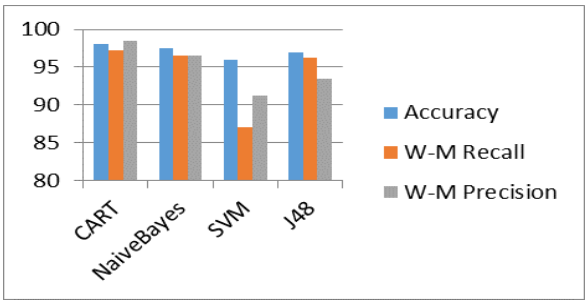


Fig. 5 R-Performance measures

V. CONCLUSION

This paper has used a machine learning algorithm that is a Decision tree algorithm for classifying the weather parameters such as Minimum Temperature, wind speed, and seasonal rainfall in terms of the seasonal month for the Coimbatore district. The result shows how these parameters have influenced the weather-based insurance factors over the study period with different data mining tools. Decision trees prove an effective decision support system in crop insurance prediction. The variation in weather conditions in 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] Report on impact evaluation of pilot weather-based crop insurance study (wbcis**),** Submitted to Department of Agriculture and Cooperation, Ministry of Agriculture, Government of India By Agricultural Finance Corporation Ltd., Head Office, Mumbai, JANUARY,2011

[3] Prof. Nilima Patil,,Prof. RekhaLathi, Prof. VidyaChitre, Comparison of C5.0& CART Classification algorithms using pruning technique, International Journal of Engineering Research & Technology (IJERT)Vol. 1 Issue 4, June - 2012

[4] Fahad Sheikh, S. Karthick1 , D. Malathi, J. S. Sudarsan, and C. Arun1, Analysis of Data Mining Techniques for Weather Prediction,, Indian Journal of Science and Technology, Vol 9(38), DOI: 10.17485/ijst/2016/v9i38/101962, October 2016

[5] Siddharth S. Bhatkande1, Roopa G. Hubballi2, Weather Prediction Based on Decision Tree Algorithm Using Data Mining Techniques, IJARCCE,2016

[6] ZaheerUllah 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.

[7] https://www.wisdomjobs.com/e-university/r-programming-language-tutorial-1579/r-decision-tree-18325.html

[8] Rajesh Kumar, Decision tree for the weather forecasting, International Journal of Computer Applications (0975 – 8887) vol.2, August 2013.

[9] Annie Mary Bhavitha S, Sudha Madhuri, A Classification Method using Decision Tree for Uncertain Data, International Journal of Computer Trends and Technology, Volume-3 Issue-1,2012