ISSN:2348-2079

Volume-6 Issue-2

International Journal of Intellectual Advancements and Research in Engineering Computations

# Crop Yield Prediction in Tamil Nadu using Baysian Network

[1]Mrs. K. E. Eswari, M.C.A.,M.Phil., M.E., Associate Professor/MCA
[2]Ms. L.Vinitha, III MCA
Department of MCA,Nandha Engineering College(Autonomous),Erode-52
E-Mail ID: eswari.eswarimoorthy@nandhaengg.org,Vinithalakshmi95@gmail.com

***Abstract*-Crop prediction is an important agricultural problem. To address this problem, clustering and classification techniques are used for crop yield prediction. It is the one of the most commonly used intelligent technique based on data analytics concepts to predict the crop yield for maximizing the crop productivity. Machine learning techniques can be used to improve prediction of crop yield under different climatic scenarios. The Baysian network Classification is a supervised learning model which means temperature and rainfall analyzes the crop data used for classification and probability values of Rice, Coconunt, Arecanut, Black pepper and Dry ginger crops. .The Baysian network Classification analysis technique is used for exploring the dataset.For the present study the mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE) and root relative squared error (RRSE) were calculated. The experimental results showed that the performance of other techniques on the same dataset was much better compared to SMO.**

**Index Terms- Agriculture, Bayesian network, crop, parameter, dataset**

## I.INTRODUCTION

Farmers are faced with having to make difficult decisions as to how to remain productive and sustainable with changing climates and market economic pressure. The provision of accurate and timely information such as meteorological, soil, use offertilizers, use of pesticides can assist farmers to make the best decision for their cropping situations.This could benefit them to attain greater crop productivity if the conditions are suitable or help them to decrease the loss due to unsuitable conditions for the crop yield. A number of studies have investigated how Information and Communication Technologies (ICT) can be applied to improve crop yield prediction and have successfully implemented in various climatic scenarios [1,2,3,4,5,6,7,8,].In addition proposed to that, plot each data item as a

point in n-dimensional spacewith the value of each feature being the value of a particular coordinate.

*S.Veenadhari et al[10] [2011]*describe an attempt has been made to review the research studies on application ofdata mining techniques in the field agriculture. Some of the techniques, such asID3 algorithms, the k-means, the k nearestneighbor, artificial neural networks and support vectormachines applied in the field of agriculture were presented. Data mining in application in agriculture is a relatively new approach for forecasting / predicting of agricultural crop/animal management.

*A Gonzalez Sanchez et al [11] [2014]* describe an important issue for agricultural planning purposes is the accurate yield estimation for the numerous crops involved in the planning. Machine learning (ML) is an essential approach for achieving practical and effective solutions for this problem. Many comparisons of ML methods for yield prediction have been made, seeking for the most accurate technique. Generally, the number of evaluated crops and techniques is too low and does not provide enough information for agricultural planning purposes. This paper compares the predictive accuracy of ML and linear regression techniques for crop yield prediction in ten crop datasets.

## II.RELATED WORK

Support Vector Machines (SVMs) a supervised machine learning technique. There are a number of examples of where it has been used in the agricultural domain. Tripathi*et al.*, (2006) reported on how SVM was applied for reduction of precipitation for climate change scenarios [9]. To minimize the generalization error bound and to achieve generalized performance, SVM was used to forecast the demand and supply of pulp wood [10].SVM was also applied to provide insights into crop response patterns related to climate conditions by providing the features contribution analysis for agricultural yield prediction [11].

## III.RESEARCH METHODS

This section discusses the methods used for this research and includes details of the study area, datasets and methodology.

### *A.Study Area*

The agricultural dataset is submitted to the system for the crop prediction and classification process. From the vast initial dataset, only a limited number of important factors which have the highest impact on agricultural yield were selected for the present research

Maharashtra has a long coastline stretching nearly 720 kilometers along the Arabian Sea and occupies the western and central part of the country . Figure 1 below shows the study area selected for this research. The state has a geographical area of 3,07,713 sq. km. It is bounded by North latitude 15°40' and 22°00' and East Longitudes 72°30' and 80°30'.

The state has 35 districts which are divided into six revenue divisions viz. Konkan, Pune, Nashik, Aurangabad, Amravati and Nagpur for administrative purposes. For the present research, 27 districts were selected as representatives of the state depending on the data availability. The states selected were Ahmednagar, Amravati, Aurangabad, Beed/Bid, Bhandara, Buldhana, Chandrapur, Dhule, Gadchiroli, Gondia, Hingoli, Jalana, Jalgaon, Kolhapur, Latur, Nagpur, Nanded, Nasik, Osmanabad, Parbhani, Pune, Sangli, Satara, Solapur, Wardha, Washim and Yavatmal. Principal crops grown in the state are rice, jowar, bajra, wheat, tur, mung, urad, gram and other pulses .

### *B. Dataset Used*

All the datasets used in the research were sourced from the openly accessible records of the Indian Government. This was sourced for the years 1998 to 2002 for the Kharif season of rice production. From the vast initial dataset, only a limited number of important factors which have the highest impact on agricultural yield were selected for the present research.

- Precipitation (mm): The total precipitation for Kharif season (June to November) for each year of every district was calculated from the monthly mean precipitation of that year for a particular district.

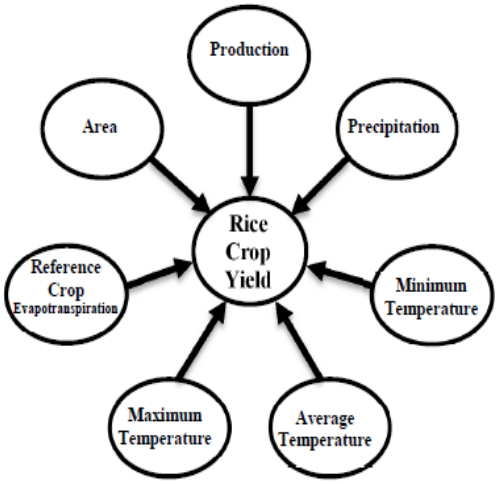


Figure 2 Study of climatic parameters on Maharashtra State, India

- Minimum, Average, Maximum Temperature (degree Celsius): Crop production will definitely have an impact due to variation in the temperature. Hence minimum, average and maximum temperature for each year of every district was considered for the present research. The average temperatures for the Kharif season (June to November) were calculated from the monthly mean temperature for minimum, average and maximum temperatures of that year for every district.

- Reference Crop Evapotranspiration (mm): The reference crop evapotranspiration was calculated on the basis of monthly mean of that year for the Kharif season for every district.

- Area (Hectares): The rice cultivated area in Kharif season (June to November) for every year in each selected district of Maharashtra state was considered for the present research.

*C. Methodology Used*

The following steps were followed to prepare the data for processing after incorporating all the datasets of this study in to Microsoft Office Excel.

**Step 1**: Acquiring each parameter (precipitation, minimum, average, maximum temperature and reference crop evapotranspiration) monthly mean records of each district from 1998 to 2002 from the Indian Government records.

**Step 2**: Calculating the total precipitation, average temperature for the minimum, average and maximum temperature and average reference crop evapotranspiration for each year for each district during the Kharif season (June to November) of Maharashtra state.

**Step 3:** Acquiring each districts area, production and rice crop yield details of the year 1998 to 2002 from the publicly available Indian Government records.

**Step 4**: The raw data set was then collated in single sheet which consisted of the following columns in Microsoft Excel: sr. no, name of the district, year, precipitation, minimum temperature, average temperature, maximum temperature, reference crop evapotranspiration, area, production and yield.

**Step 5**: For some of the districts particular year's climatic parameters or production data was not available hence those year's data was not used for the current research. Record number was added for each record.

**Step 6**: For preparing the data set for applying data mining techniques, unrequired columns were omitted. They were sr. no, name of the district and year.

**Step 7**: The data set was then sorted on the basis of area. Area less than 100 hectares were not considered for the present research. So those records were omitted.

**Step 8**: The dataset was then sorted on the basis of yield to classify the records in to low, moderate and high. The low yield was from 0.15 to 0.60 tonnes/hectare, moderate from 0.61 to 1.10 tonnes/hectare and high from 1.11 to 3.16 tonnes/hectare. Class low had 45 records with the range 0.15 to 0.60 tonnes/hectare, class moderate had 46 records with the range 0.61 to 1.10 tonnes/hectare and class high had 44 records with the range 1.11 to 3.16 tonnes/hectare.

**Step 9**: The yield has been calculated on the basis of area and production hence these two columns were omitted.

**Step 10**: This data set was then saved in .csv format for further applying data mining techniques. This file had following columns: precipitation, minimum temperature, average temperature, maximum temperature, reference crop evapotranspiration, crop yield and class.

The WEKA tool is a freely available and open source data mining tool available under the GNU General Public License.The data set prepared for the present exploration was saved in .arff file format and processed through WEKA to build the algorithm on the current data set.

Support Vector Machine

Crop mapping is widely used in agriculture and remote sensing science. Crop aping using classification methodologies serves various applications in agricultural science like gauging water and temperature demand etc. For such applications information on the spatial distribution of classification error is of particular interest. Recent progresses in Information Technology systems, lead to dynamic communication among user of every profession.

- Acquiring each parameter (precipitation, minimum, average, maximum temperature and reference crop evapotranspiration) monthly mean records of each district every year from the Indian Government records.
- Calculating the total precipitation, average temperature for the minimum, average and maximum temperature and average reference crop evapotranspiration for each year for each district during the season form June to November of Tamilnadu state.

The SVM approach [17] is used to create functions from a set of labeled training data. These functions can be a classification function or it can be general regression function. For the current study SMO algorithm was used to study the performance of this approach on the dataset used for the present study.

## VI.BAYESIAN NETWORK CLASSIFIERS

Bayesian network classifiers are used in many rice crop fields and one common class of classifiers are naive Bayes classifiers. In this thesis, we introduce an approach for reasoning about Bayesian network

classifiers in which we explicitly convert them into Ordered Decision Diagrams (ODDs), which are then used to reason about the crop properties of these classifiers. Specifically, we present an algorithm for converting any naive Bayes classifier into an ODD, and we show simulation that this algorithm can give us an ODD that is tractable in size even given an intractable number of instances. Numerous approaches to this problem are based on various functional representations such as decision trees, decision lists, neural networks, decision graphs, and rules.given*C*whenever Pr(*A*|*B*, *C*) = Pr(*A*|*C*) for all possible values of *A*, *B* and *C*, whenever Pr(*C*) >0.

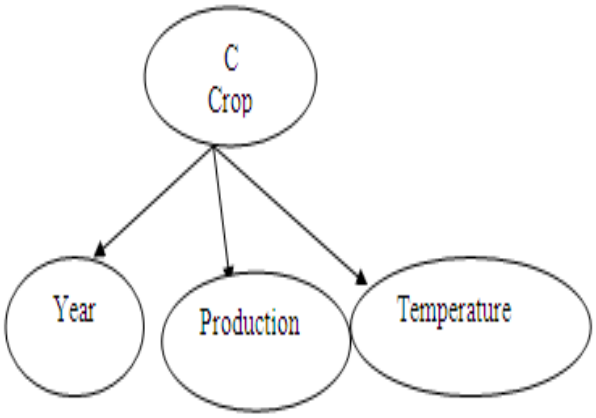


Fig 3.3.1 Structure of the Naive Bayes Network

Bayesian network B, that encodes a distribution PB (Crop Year, Crop Production, Crop Temperature, Crop_Rainfail, Crop_n), from a given training set. We can then use the resulting model so that given a set of attributes a1. . . an, the classifier based on B returns the label c that maximizes the posterior probability PB(c|Crop_Yea_a1, Crop_Production_a2, . . . , Crop_an). Note that, by inducing classifiers in this manner, we are addressing the main concern expressed in the introduction: remove the bias by the independence assumptions embedded in the naïve Bayesian classifier.this manner, we are addressing the main concern expressed in the introduction: remove the bias by the independence assumptions embedded in the naïve Bayesian classifier.

To understand the possible discrepancy between good predictive accuracy and good MDL score, we must re-examine the MDL score. Recall that the log likelihood term in Equation 2 is the one that measures the quality of the learned model, and that D = {Crop_u1,. . ,Crop_uN} denotes the training set. In a classification task, each Crop_uiis a tuple of the form Crop_hai1, . . . ,Crop_ ain,Crop_ cii that assigns values to the attributes Crop_A1, . . . ,Crop_ An and to the class variable Crop_C. We can rewrite the log likelihood function as

LL(Crop_B|Crop_D) = N Xi =1 log PB(Crop_ci|Crop_ai1, . . . ,Crop_ ain) + N Xi =1 log PB(Crop_ai1, . . . , Crop_ain)

BAYESIAN NETWORKS ALGORITHM

- Step1: Read the Crop dataset.
- Step 2: Create the data list from Crop dataset and feature extraction for prediction crop details.
- Step 3: Create the Bayesian net using Neuralnet package
- Step 4: To create a Rice, Coconunt, Arecanut, Black pepper and Dry ginger crop base net using model2network function in R-Studio.
- Step 5: To read the dataset and assign the data into CA, CS, Ck, Cw, CB, CL and CE object variable. The object contains Crop year, Crop production, Crop Area, Crop mean Temperature, Crop mean Rainfall, mean Crop Temperature and Rainfall values, and districts details is connected the Bayesian network.
- Step 6: The crop classification rule and probability values assign the Bayesian net.
- Step 7: To create custom Bayesian net using Bayesian theory in Rice, Coconunt, Arecanut, Black pepper and Dry ginger crop.
- Step 8: To check the Bayesian Rule for Rice, Coconunt, Arecanut, Black pepper and Dry ginger crop and return the accuracy values.
- Step 9: Repeat the process Step 3 to Step 8.
- Step 10: To accuracy calculate the TP, TN, FP and FN values.

## V. PERFORMANCE EVALUATION

Each instance is classified into two classes in a binary classification model. The two classes are true and false class. This gives rise to four possible classifications for each instance namely:

True Positive (TP): The number of correct predictions that an instance is positive.
False Positive (FP): The number of incorrect predictions that an instance is positive.
False Negative (FN): The number of incorrect predictions that an instance is negative.
True Negative (TN): The number of correct predictions that an instance is negative.

This situation can be depicted as a confusion matrix also called contingency table as shown in table 1 below.

| | | Observed | |
|---|---|---|---|
| | | True | False |
| Predicted | True | True Positive (TP) | False Positive (FP) |
| | False | False Negative (FN) | True Negative (TN) |

The observed classifications for a phenomenon are compared with the predicted classifications of a model in a confusion matrix. In table 1 the classification that are shown along the
major diagonal of the table are the correct classifications refereed as true positives and true negatives. The model errors are signified by the other fields. Only the true positive and true negative fields would be filled out for a perfect model and the other fields would be set to zero. From the confusion matrix, a number of model performance metrics can be derived.The most common metric is accuracy which is defined as the overall success rate of the classifier and is computed as

$$\text{Accuracy} = (TP + TN) / (TP + FP + FN + TN)$$

Other performance metrics include Sensitivity/Recall and Specificity/Precision. Sensitivity is defined as percentage of correctly classified instances. Specificity is defined as percentage of incorrectly classified instances. These can be computed as

$$\text{Sensitivity/Recall} = TP / (TP + FN)$$

$$\text{Specificity/Precision} = TP / (TP + FP)$$

F1 Score is a measure of test's accuracy. To compute the score it considers precision and recall. F1 score is the harmonic mean of precision and recall. F1 score can be computed as

$$F1 = (2TP)/(2TP + FP + FN)$$

The weighted average of the precision and recall is referred as F1 score. It reaches its best value at 1 and worst at 0.

Mathews Correlation Coefficient (MCC) is computed as a measure of the quality of classification. It is considered as a balanced measure that can be used even if the classes are of different sizes by considering true and false positives and negatives.

The MCC returns a value between −1 and +1 which is a correlation coefficient between the observed and predicted binary classifications. A perfect prediction is represented by coefficient of +1, random prediction by 0 and total disagreement between prediction and observation by -1. It can be computed as shown below.

$$MCC=((TP*TN)-(FP*FN))/\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}$$

For reference and evaluation, the relative mean absolute error (MAE), root mean squared error (RMSE), relative absolute error (RAE) and root relative squared error (RRSE) are also computed.

## VI.CONCLUSION

In recent years, great efforts have been undertaken on the challenging task of predicting rice crop yield. Developing accurate models for crop yield estimation using Information and Communication Technologies may help farmers and other stakeholders improve decision making in relation to national food import/exports and food security. Rice is one of the most important food crops of India.

It is cultivated all over the country and contributes more than 40% of total food grain production [9]. Given the importance of rice to world's food security, any improvements in the forecasting of rice crop yield under different climatic and cropping scenarios will be beneficial. This research has demonstrated the

prediction of rice crop yield by applying one of the machine learning technique, support vector machine (SVM).

The experimental results showed that the other classifiers such as Naïve Bayes, BayesNet and Multilayer Perceptron performed better by achieving the highest accuracy, sensitivity and specificity compared to SMO classifier with lowest accuracy, sensitivity and specificity that has been reported earlier for the same data set [10,11].

In terms of test's accuracy and quality also BayesNet and Multilayer Perception showed the highest accuracy and best quality and SMO showed the lowest accuracy and worst quality. It can be concluded that other classifiers used on the current study dataset and reported earlier should be recommended for further development of a rice prediction model .

## REFERENCES

[1] R. Medar, V. Rajpurohit, "A survey on data mining techniques for crop yield prediction", International Journal of Advance Research in Computer Science and Management Studies, vol. 2, no. 9, pp. 59-64, 2014.

[2] S. Bejo, S. Mustaffha and W. Ismail, "Application of artificial neural network in predicting crop yield: A review", Journal of Food Science and Engineering, vol. 4, pp.1-9, 2014.

[3] S. Dahikar and S. Rode, "Agricultural crop yield prediction using artificial neural network approach", International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, vol. 2, no. 1, pp. 683-686, 2014.

[4] W. Guo and H. Xue, "An incorporative statistic and neural approach for crop yield modelling and forecasting", Neural Computing and Applications, vol. 21, pp. 109-117, 2012.

[5] W. Guo and H. Xue, "Crop yield forecasting using artificial neural networks: A comparison between spatial and temporal models", Mathematical Problems in Engineering, pp.1-7, 2014.

[6] D. Ramesh and B. Vardhan, "Analysis of crop yield prediction using data mining techniques", International Journal of Research in Engineering and Technology, vol. 4, no. 1, pp. 47-473, 2015.

[7] K. Tanaka and T. Kiura, "Crop yield prediction systems for rainfed areas and mountainous areas in Thailand", Proceedings of the 9th Conference of the Asian Federation for Information Technology in Agriculture "ICT's for future Economic and Sustainable Agricultural Systems", 2014.

[8] G. Yengoh and J. Ardo, "Crop yield gaps in Cameroon", AMBIO, Springer, vol. 43, pp. 175-190, 2014.

[9] Tripathi, V.V. Srinivas and R.S.Nanjundiah, "Downscaling of precipitation for climate change scenarios: a support vector machine approach", Journal of Hydrology, vol. 330, no. 3, pp.621-640, 2006

[10] Veenadhari S, Misra B, Singh CD. Data mining techniques for predicting crop productivity—A review article. In: IJCST. 2011; 2 (1).

[11] Gonzalez-Sanchez Alberto, Frausto-Solis Juan, Ojeda-Bustamante W. Predictive ability of machine learning methods for massive crop yield prediction. Span J Agric Res.2014;12(2):313–2.