Crop Yield Prediction using Machine Learning Techniques

Ramesh Medar
KLS GIT, Belagavi
rameshcs.git@gmail.com

Vijay S. Rajpurohit
KLS GIT, Belagavi
vijaysr2k@yahoo.com

Shweta
KLS GIT, Belagavi
shwetupeerapur@gmail.com

Abstract: **Agriculture is the field which plays an important role in improving our countries economy. Agriculture is the one which gave birth to civilization. India is an agrarian country and its economy largely based upon crop productivity. Hence we can say that agriculture can be backbone of all business in our country. Selecting of every crop is very important in the agriculture planning. The selection of crops will depend upon the different parameters such as market price, production rate and the different government policies. Many changes are required in the agriculture field to improve changes in our Indian economy. We can improve agriculture by using machine learning techniques which are applied easily on farming sector. Along with all advances in the machines and technologies used in farming, useful and accurate information about different matters also plays a significant role in it. The concept of this paper is to implement the crop selection method so that this method helps in solving many agriculture and farmers problems. This improves our Indian economy by maximizing the yield rate of crop production.**

Keywords: Indian Agriculture, Machine Learning Techniques, Crop selection method.

I. INTRODUCTION

The main goal of agricultural planning is to achieve maximum yield rate of crops by using limited number of land resources. Many machine learning algorithms can help in improving the production of crop yield rate. Whenever there is loss in unfavourable conditions we can apply crop selecting method and reduce the losses. And it can be used to gain crop yield rate in favourable conditions. This maximising of yield rate helps in improving countries economy. We have some of the factors that influence the crop yield rate. They are seed quality and crop selection. We need test the quality of the seeds before sowing. As we know that good quality of seeds helps in getting more yield rate. And selection of crops depends upon two things that is favourable and unfavourable conditions. This can also be improved by using hybridization methods. Many researches are carried out to improve agricultural planning. The goal is to get the maximum yield of crops. Many classification methods are also applied to get maximum yield of crops. Machine learning techniques can be used to improve the yield rate of crops. The method of crop selection is applied to improve crop production.

The production of crops may depend on geographical conditions of the region like river ground, hill areas or the depth areas. Weather conditions like humidity, rainfall, temperature, cloud. Soil type may be clay, sandy, saline or peaty. Soil composition can be copper, potassium, phosphate, nitrogen, manganese, iron, calcium, ph value or carbon and different methods of harvesting. Many parameters are used for different crops to do different predictions. These prediction models can be studied by using researches. These predictions are classified as two types. One is traditional statistic method and other is machine learning techniques. Traditional method helps in predicting single sample spaces. And machine learning methods helps in predicting multiple predictions. We need not to consider the structure of data models in traditional method where as we need to consider the structure of data models in machine learning methods.

II. LITERATURE SURVEY

In [1] **J.P. Singh, Rakesh Kumar, M.P. Singh and Prabhat Kumar**, have concluded that this paper helps in improving the yield rate of crops by applying classification methods and comparing the parameters. We can also do analysing and prediction of crops using baysian algorithms. The algorithms used are Bayesian algorithm, K-means Algorithm, Clustering Algorithm, Support Vector Machine. The disadvantage is that there is no proper accuracy and performance.

In [2] the authors **Subhadra Mishra, Debahuti Mishra and Gour Hari Santra**, have concluded that this is an advanced researched field and is expected to grow in the future. The integration of computer science with agriculture helps in forecasting agricultural crops. This method also helps in providing information of crops and how to increase yield rate. The algorithms used are Artificial neural networks, Decision Tree Algorithms, Regression analysis. The disadvantage is clear methodology is not specified.

In [3] the authors **Karan deep Kauri**, have concluded that this paper will review that various applications of machine learning in the farming sector. And also provides an insight into the troubles faced by our Indian farmers and how these can be solved using these techniques. This method help in increasing the farming sector in the countries and apply the more machine learning applications. The algorithms used are Artificial neural networks, Bayesian Belief Network, Decision Tree Algorithms, Clustering, Regression analysis. The disadvantage is less accuracy in terms of performance.

In [4] **E. Manjula, S. Djodiltachoumy**, have concluded that the aim of this paper is to propose and implement a rule based system. And predict the crop yield production from the collection of previous data. The algorithms used are K-means Algorithm, clustering method. The disadvantage is Suitable only for using association rule and considered less data.

In [5] **Nishit Jain, Amit Kumar, Sahil Garud, Vishal Pradhan, Prajakta Kulkarni**, have concluded that this paper helps in predicting crop sequences and maximizing yield rates and making benefits to the farmers. Also, Using machine learning applications with agriculture in predicting crop diseases, studying crop simulations, different irrigation patterns. The algorithms used are Artificial neural networks, Support Vector Machine. The disadvantage is Exact accuracy is not specified.

In [6] **B.Mallikarjun Rao, D.Sindhura, B.Navya Krishna, K.Sai Prasanna Lakshmi, Dr. J Rajendra Prasad**, have concluded that this method will provide a useful and accurate knowledge. Using this knowledge we predict and support the decision making for different sectors. The algorithms used are multiple linear regressions. The disadvantage is that it can be applied for limited areas.

In [7] **T.Giri Babu, Dr.G.Anjan Babu,** have concluded that this method will provide solutions to the farmers. They can also help in providing solution for water and fertilizer problems. And this helps to get more production of yield. The algorithms used are agro algorithm. The disadvantage is that this method does not give proper accuracy for crops.

In [8] **B Vishnu Vardhan, D Ramesh,** have concluded that this method will provide multiple linear regression method which can be applied for existing data and hence helps in analyzing and verifying the data. The algorithms used are multiple linear regressions. The disadvantage is that it results in less accuracy.

In [9] **Ashwani Kumar Kushwaha, SwetaBhattachrya ,** have concluded that this method will provide agro algorithm which helps in predicting suitable crop for the lands. And this helps in enhancing the quality of crop. The algorithm used is agro algorithm. The disadvantage is it results in less prediction of crops.

In [10] **Raorane A.A, Dr. Kulkarni R.V,** have concluded that this method will help in estimating rain fall and investigate the reasons for getting lower yield. The algorithm used is regression analysis method. The disadvantage is that here the specific method is not specified.

In[11] **Anshal Savla, Himtanaya Bhadada, Vatsa Joshi, Parul Dhawan ,** have concluded that this method will help in analyzing and understanding crop yield rate for zones which is based on attributes. The algorithms used are Normalization, Clustering, and Classification. The disadvantage is it gives only framework.

In [12] **Siti Khairunniza-Bejo, Samihah Mustaffha, Wan Ishak Wan Ismail ,** have concluded that this method will help in giving solutions to the few problems of farmers in getting good yield. The algorithms used are Artificial Neural Network. The disadvantage is it consumes more time.

III. IMPLEMENTATION

Here we are going to use two different methods. First is Naive Bayes method and second is K-Nearest neighbour method. We can get the accuracy of performance by using these two methods. To predict the crop yield rate a java application is created. This application includes three parts. First is managing datasets second is testing datasets and third is analyzing the datasets. In managing datasets we can get the datasets of previous years and they can also be converted into supporting format. As we are using Weka tool in this project all the datasets are converted to attribute relation file format. In testing part we can do the single testing. We have considered two methods of machine learning. One is Naive Bayes and other is K-Nearest neighbour method. In testing we can select any one of the method and do testing of dataset like by selecting particular crop, particular place and particular season we can get results of yield. In analyzing part we can input a whole dataset file and get accuracy of the two different methods. This helps in predicting which method is good.

As the farmers are facing many problems in agricultural sector we need to minimize their problems. This reduction of problems can be done by implementing new techniques on agriculture. We can implement the machine learning methods on agriculture. We have clustering and classification methods that can be applied on crops. We can also apply some of regression methods to improve the production of crops. In this project we have considered only the Naive Bayes method and K-Nearest Neighbour method. Using these two methods we can predict which crops to be selected for their land and season. As the farmers don't know to use Weka tool we have developed a java application. This application helps them to predict the yield.

Here in this application we can do single testing by giving input as crop name, season selected and place selected. We can use any method among KNN or NB method. As soon you give the input you can select the method and mine the results. The results will tell you the yield rate of that crop. And we can do multiple testing by analyzing the datasets. In analyzing it allows you to select a whole file at once and get the accuracy. Here instead of keep on doing single tests we can directly do the multiple testing. This testing helps in getting the accuracy between two methods. By this we will come to know which method is good among given methods. And this will help the farmers which crop to be selected for their land or the region. The datasets include the results of previous year data. These datasets help in predicting the results for new instances. Farmers can give any instance to the test and get the yield rate for the crop. So this application helps farmers to select the proper crop for land. And it also helps them to predict the yield rate of selected crop. These methods can be implemented manually. Here we consider the probability values of instances. We can get the result for new instances. The Naive Bayes method will find the probability of good and poor. And we can predict whether the crop selected will give good yield or poor yield rate. Similarly the KNN method will calculate the distance between two values given to the instances and finds the minimum value. This method uses Euclidian distance to get the distance between two values.

A. N aive Bayes Algorithm
Naive bayes algorithm is completely based on naive bayes classifier. This classifier helps in finding the probability of predicted classes. This method is easy in building large datasets.

$$P(C \mid X) = \frac{P(X \mid C) * P(C)}{P(X)}$$

Bayes theorem allows to calculate posterior probability P(C|X) from the given P(X|C), P(X) and P(C).

P(C|X) = conditional probability of X when given C that is the posterior probability.

P(X|C) = conditional probability of C when given X that is likelihood. P(C) = prior probability of C.

P(X) = probability of X

Naive Bayes Classifier

$$P(a_i \mid v_j) = \frac{n_c + mp}{n + m}$$

- n is number of training examples for which $v=v_j$.
- nc is number of examples for which $v=v_j$ and $a=a_i$.
- p is prior estimation for $P(a_i \mid v_j)$.
- m is equivalent sample size. Example:

TABLE 1: DATASETS FOR CROP YIELD PREDICTION

Example No	Crop	District	Season	Yield
1	Rice	Belgaum	Kharif	Good
2	Rice	Belgaum	Kharif	Poor
3	Rice	Belgaum	Kharif	Good
4	Wheat	Belgaum	Kharif	Poor
5	Wheat	Belgaum	Rabi	Good
6	Wheat	Bijapur	Rabi	Poor
7	Wheat	Bijapur	Rabi	Good
8	Wheat	Bijapur	Rabi	Poor
9	Rice	Bijapur	Rabi	Poor
10	Rice	Hubli	Rabi	Good

Training example
We are going to classify Rice Kharif Bijapur. And there is no example of Rice Kharif Bijapur in the datasets. Lets calculate the probabilities.

P(Rice | Good) , P(Bijapur | Good) , P(Kharif | Good) ,

P(Rice | Poor) , P(Bijapur | Poor) and P(Kharif | Poor) .

Now multiply both of them by P(Good) and P(Poor). We will estimate the values.

Good:
Rice: n = 5, n_c = 3, m = 0.5, p = 3.
Bijapur: n = 5, n_c = 1, m = 0.5, p = 3.
Kharif: n = 5, n_c = 2, m = 0.5, p = 3.

Poor:
Rice: n = 5, n_c = 2, m = 0.5, p = 3.
Bijapur: n = 5, n_c = 3, m = 0.5, p = 3.
Kharif: n = 5, n_c = 3, m = 0.5, p = 3.

P(Rice | Good) = 3+3*0.5 / (3+5) =0.56 P(Rice | Poor) = 2+3*0.5 / (3+5) =0.43

P(Bijapur | Good) = 1+3*0.5 / (3+5) =0.31 P(Bijapur | Poor) = 3+3*0.5 / (3+5) =0.56

P(Kharif | Good) = 2+3*0.5 / (3+5) =0.43 P(Kharif | Poor) = 3+3*0.5 / (3+5) =0.56

We have P(Good) = 0.5 and P(Poor) = 0.5, so we can now apply equation (2). For v = Good, we have

P(Good) * P(Rice | Good) * P(Bijapur | Good) * P(Kharif |Good)
= 0 .5 * 0.56 *0 .31 *0 .43 = 0.037 and for v = No, we have

P(Poor) * P(Rice | Poor) * P(Bijapur | Poor) * P(Kharif | Poor)
= 0.5 * 0.43 * 0.56 * 0.56 = 0.069 Since

0.069 > 0.037, our example can be classified as "POOR".

B. KNN (K- Nearest Neighbor)
K-nearest neighbor method can be used for both regression and classification predictive problems. This method helps in interpret output, calculate time and predictive power. The Machine learning techniques are used in various fields. KNN is also one of the machine learning method. This is also called as method of sample based learning. This will contain the data of past datasets and can be used while predicting the new datasets. This will apply function called as distance function like Manhattan or Euclidean distance. This can be used to compute distance from samples to all other training samples. It calculates the target value for new samples. The target vale will be the weighted sum of target values of the k nearest neighbours. The valve of K can be directly proportional to the prediction. Whenever the valve of K is small this indicates there is high variance and there is low bias. If the valve of the K is larger than this indicates that there is low variance and high bias. The main advantage of this KNN is it does not require any training or the optimization. This KNN uses data samples when predicting the new datasets. Hence it is having higher complexity and also more time consumption.
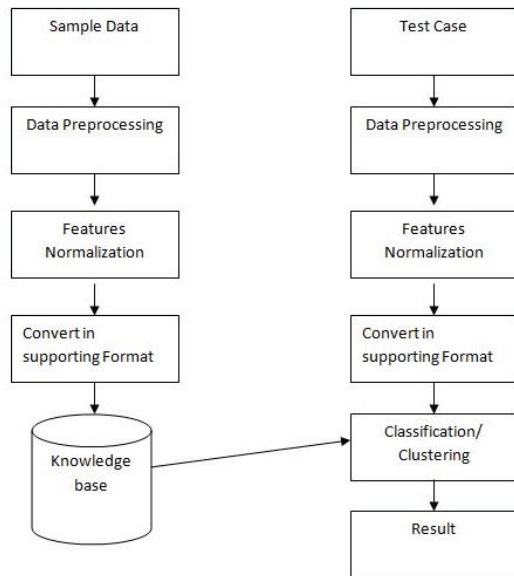
IV. SYSTEM ARCHITECTURE

Fig. 1: System Architecture for Crop Yield Prediction

Here first we collect the data sets and process the data and we remove if there are any impurities in the data sets. Next the data is normalized if needed like it can be converted to smaller volume of data. Next the data is converted to supporting format. And then it is stored in the databases. Next the required method is applied. Now we get the final results.

V. RESULTS

To predict the crop yield rate a java application is created. This application includes three parts. First is managing datasets second is testing datasets and third is analyzing the datasets. In managing datasets we can get the datasets of previous years and they can also be converted into supporting format. As we are using Weka tool in this project all the datasets are converted to attribute relation file format. In testing part we can do the single testing. We have considered two methods of machine learning. One is Naive bayes and other is K-Nearest neighbour method. In testing we can select any one of the method and do testing of datasets like by selecting particular crop, particular place and particular season we can get results of yield. In analyzing part we can input a whole dataset file and get accuracy of the two different methods. This helps in predicting which method is good.

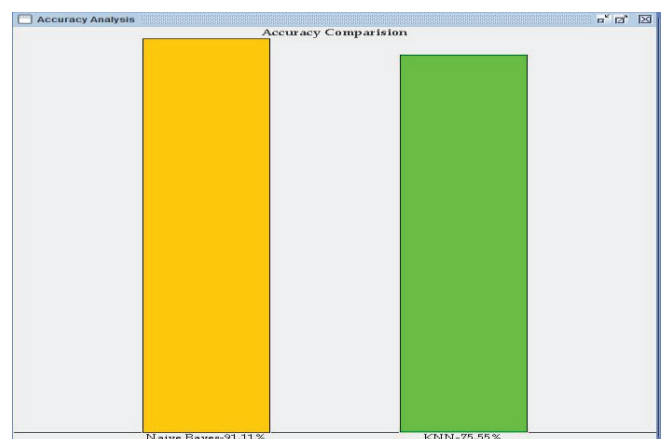Fig. 2: Original datasets

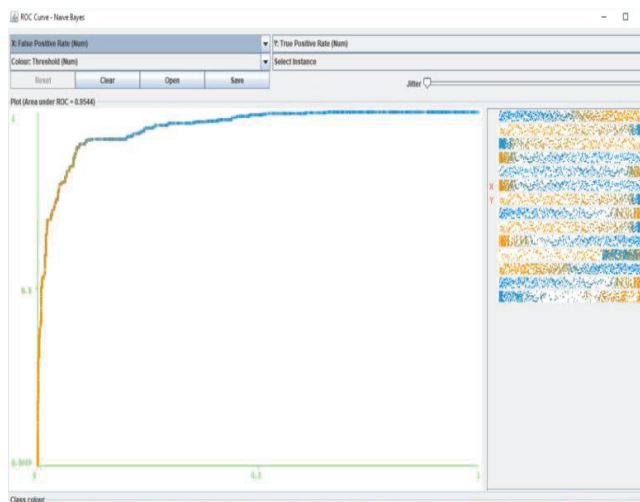Fig. 3: New datasets

Fig. 4: Result of checking accuracy

Fig. 5: ROC curve for the results

VI. CONCLUSION

Agriculture is the field which helps in economic growth of our country. But this is lacking behind in using new technologies of machine learning. Hence our farmers should know all the new technologies of machine learning and other new techniques. These techniques help in getting maximum yield of crops. Many techniques of machine learning are applied on agriculture to improve yield rate of crops. These techniques also help in solving problems of agriculture. We can also get the accuracy of yield by checking for different methods. Hence we can improve the performance by checking the accuracy between different crops. Sensor technologies are implemented in many farming sectors. This paper helps in getting maximum yield rate of the crops. Also helps in selecting proper crop for their selected land and selected season. These techniques will solve the problems of farmers in agriculture field. This will help in improving the economic growth of our country.

REFERENCES

[1] J.P. Singh, M.P. Singh, Rakesh Kumar and Prabhat Kumar Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique, International Journal on Engineering Technology, May 2015.

[2] Gour Hari Santra, Debahuti Mishra and Subhadra Mishra, Applications of Machine Learning Techniques in Agricultural Crop Production, Indian Journal of Science and Technology, October 2016.

[3] Karan deep Kauri, Machine Learning: Applications in Indian Agriculture, International Journal of Advanced Research in Computer and Communication Engineering, April 2016.

[4] S. Djodiltachoumy, A Model for Prediction of Crop Yield, International Journal of Computational Intelligence and Informatics, March 2017.

[5] Nishit Jain, Amit Kumar, Sahil Garud, Vishal Pradhan, Prajakta Kulkarni, Crop Selection Method Based on Various Environmental Factors Using Machine Learning, Feb -2017.

[6] D.Sindhura, B.Navya Krishna, K.Sai Prasanna Lakshmi, B.Mallikarjun Rao, Dr. J Rajendra Prasad, Effects of Climate Changes on Agriculture International Journal of Advanced Research in Computer Science and Software Engineering, March 2016.

[7] T.Giri Babu, Dr.G.Anjan Babu, Big Data Analytics to Produce Big Results in the Agricultural Sector, March 2016.

[8] D Ramesh , B Vishnu Vardhan, Analysis Of Crop Yield Prediction Using Data Mining Techniques, International Journal of Research in Engineering and Technology, Jan-2015,

[9] Ashwani Kumar Kushwaha, SwetaBhattachrya, Crop yield prediction using Agro Algorithm in Hadoop, April 2015.

[10] Raorane A.A, Dr. Kulkarni R.V, Application Of Datamining Tool To Crop Management System, January 2015.

[11] Anshal Savla, Himtanaya Bhadada, Parul Dhawan, Vatsa Joshi, Application of Machine Learning Techniques for Yield Prediction on Delineated Zones in Precision Agriculture, May 2015.

[12] Siti Khairunniza-Bejo, Samihah Mustaffha and Wan Ishak Wan Ismail , Application of Artificial Neural Network in Predicting, Journal of Food Science and Engineering, January 20, 2014