www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

Machine Learning for Crop Yield Forecasting

[1] B. Naresh, [2] E. Navya,

[1]Assistant Professor, Megha Institute of Engineering & Technology for Women, Ghatkesar.
[2] MCA Student, Megha Institute of Engineering & Technology for Women, Ghatkesar.

Abstract:

One area that significantly contributes to the growth of our national economy is agriculture. Civilization was born out of agriculture. Since most of India's population lives in rural areas, the country's economy is dependent on agricultural output. So, it's safe to assume that our country's agricultural sector can support all other industries. An important part of any agricultural strategy is the selection of crops. Several factors, including market price, production rate, and various government rules, will determine the crops that are selected. If we want to see improvements in the Indian economy, we need to see a lot of reforms in the agricultural sector. The use of machine learning methods, which are well-suited to the agricultural industry, may help us enhance farming. In the same way that farming has benefited greatly from technological advancements, it has also benefited greatly from the availability of reliable information on a variety of topics. In this study, we propose a strategy for using crop selection to address several issues faced by farmers and the agricultural sector. By increasing agricultural yields to their full potential, this helps the Indian economy.

Keywords:

The use of machine learning techniques in Indian agriculture and the process of selecting crops.

I. INTRODUCTION

Maximizing crop yields while making optimal use of limited land resources is the overarching objective of agricultural planning. There are a lot of machine learning techniques that can increase agricultural yields. If crop failure occurs due to adverse weather, we may use the crop selection approach to mitigate the problem. In ideal circumstances, it may be used to increase the rate of agricultural output. Improving a country's economy is facilitated by maximizing the yield rate. Some of the variables that affect the pace of agricultural output are in our possession. Two of these factors are crop selection and seed quality. Prior to planting, we must ensure that the seeds are of high quality. We all know that a higher yield rate is possible with higher-quality seeds. Additionally, both favorable and unfavorable circumstances have a role in crop selection. Using hybridization procedures may help enhance this. The goal of better agricultural planning is the subject of several studies. Maximizing agricultural production is the objective. To get the most out of crops, a lot of categorization techniques are used as well. Crop production rates may be enhanced with the use of machine learning methods. To increase harvest yields, farmers use crop selection as a strategy. River bottoms, hilly regions, and deep places are some of the geographical factors that could affect agricultural yields. Hg, rain, temperature, and cloudiness are examples of weather conditions. Different types of soil include clay, sandy, salty, and peaty. Soil composition may vary depending on factors such as pH value, carbon, nitrogen, phosphate, potassium, iron, calcium, and various harvesting techniques. For various crops, various projections are made using a wide range of characteristics. Researchers may study these prediction models. We may divide these forecasts into two categories. One is the use of machine learning techniques, while the other is the use of classic statistical methods. The conventional approach is useful for forecasting areas with a single sample. The use of machine learning techniques also facilitates the making of many forecasts. In machine learning approaches, we must take the model structure into account, while in conventional methods, it is not necessary.

II. LITERATURE SURVEY

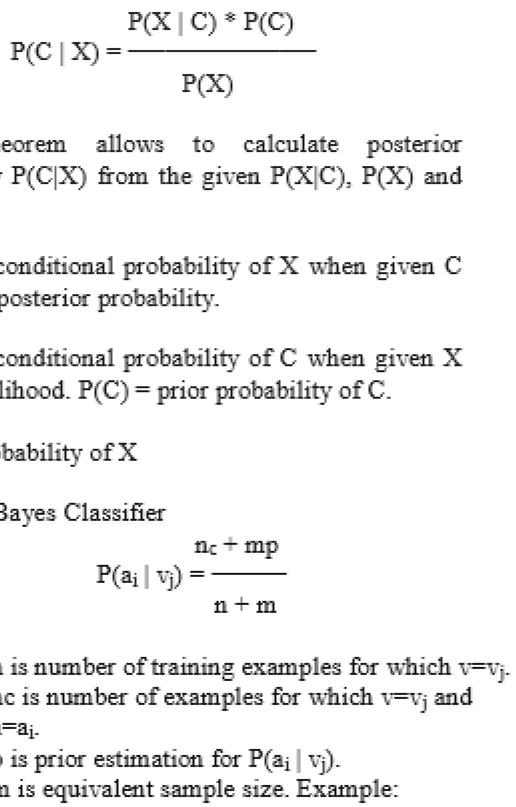
The authors of this paper—J.P. Singh, Rakesh Kumar, M.P. Singh, and Prabhat Kumar—come to the conclusion that using classification techniques and comparing the parameters helps increase agricultural production rates [1]. Using bayesian

algorithms, we can also analyze and forecast agricultural yields. The following algorithms are utilized: Bayesian, K-means, clustering, and support vector machine. The lack of sufficient precision and efficiency is a drawback. According to Subhadra Mishra, Debahuti Mishra, and Gour Hari Santra's findings in [2], this is a highly studied area that is anticipated to have future growth. Crop predictions benefit from the merging of computer science and agriculture. Information on crops and how to boost yield rate may also be provided by this way. Artificial neural networks, decision tree algorithms, and regression analysis are the algorithms that are used. The lack of clearly defined approach is a drawback. This article will examine the several uses of machine learning in agriculture, as stated by Karan deep Kauri and colleagues in [3]. Furthermore, it sheds light on the challenges encountered by our Indian farmers and offers solutions based on their specific methods. This approach facilitates the expansion of agricultural sectors in nations and the implementation of further machine learning applications. Algorithms like clustering and regression analysis, as well as artificial neural networks and Bayesian belief networks, are used. Less precision in terms of efficiency is a drawback. The purpose of this work is to suggest and execute a rule-based system, as stated in [4] by E. Manjula and S. Djodiltachoumy. And use the collected data to forecast the output of agricultural yields. The K-means algorithm and a clustering approach are used. The downside is that it is only applicable to association rules and takes into account a limited amount of data. year 2019
As stated in [5], the authors of this paper—Nishit Jain, Amit Kumar, Sahil Garud, Vishal Pradhan, and Prajakta Kulkarni—are able to assist farmers maximize their yields by accurately anticipating when crops would be planted. Additional uses for machine learning in agriculture include crop disease prediction, research into crop simulations, and exploration of alternative irrigation strategies. Support vector machines and artificial neural networks are the algorithms used. One drawback is that precise precision is not guaranteed. The authors of the cited work [6]—Dr. J. Rajendra Prasad, B. Mallikarjun Rao, D. Sindhura, B. Navya Krishna, and K. Sai Prasanna Lakshmi—have come to the conclusion that this approach would provide valuable and precise information. With this information, we can make predictions and provide assistance with sector-specific decision-making. A number of linear regression techniques are used. One drawback is that it has limited applicability. T.Giri Babu and Dr. G. Anjan Babu have come to the conclusion that this strategy will provide farmers remedies in [7]. Water and fertilizer issues may also be resolved with their assistance. More production of yield is achieved using this method. Agro algorithm is the algorithm of choice. The lack of appropriate precision for crops is a drawback of this strategy. This technique will provide a multiple linear regression approach that can be used with current data, which helps with data analysis and verification, according to B Vishnu Vardhan and D Ramesh's conclusion in [8]. A number of linear regression techniques are used. It leads to reduced precision, which is a downside. Ashwani Kumar Kushwaha and Sweta Bhattachrya have determined in [9] that this approach would provide an agro algorithm that aids in crop prediction for the land. Additionally, this contributes to an improvement in crop quality. The agro algorithm is used. The main drawback is that it leads to less accurate crop forecasting. Dr. Kulkarni R.V. and Raorane A.A. have determined that this approach will aid in rain fall estimation and investigating the causes of decreased yield in [10]. A regression analysis approach is used by the program. The lack of specificity on the procedure is a drawback. According to the findings reported in[11] by Anshal Savla, Himtanaya Bhadada, Vatsa Joshi, and Parul Dhawan, this approach will be useful for assessing and comprehending the rate of agricultural production for attribute-based zones. Classification, Normalization, and Clustering are the algorithms that are used. Constantly providing simply a framework is a drawback. According to the findings of Siti Khairunniza-Bejo, Samihah Mustaffha, and Wan Ishak Wan Ismail in [12], this approach will assist in addressing the few challenges that farmers have while trying to get a decent harvest. Artificial Neural Networks are the algorithms used. One drawback is that it takes more time.

III. IMPLEMENTATION

Two distinct approaches will be used in this case. Two methods are available: the Naive Bayes and the K-Nearest Neighbor. Using these two approaches, we can determine the performance accuracy. Predicting the rate of agricultural production requires the development of a Java program. All three components make up this application. Managing datasets comes first, followed by testing them, and then analyzing them. Datasets from prior years may be retrieved and transformed into a supported format via the management of datasets. Since this project makes use of the Weka tool, we have transformed all of the datasets into attribute relation file format. We may do individual tests in the testing phase. Two machine learning approaches have been investigated. The Naive Bayes and K-Nearest Neighbor approaches are two examples. We may test datasets using any of the available methods; for example, we can choose a certain crop, location, and season to acquire yield figures. In the analysis section, we may compare the two approaches' accuracy by inputting a whole dataset file. This is useful for determining the best approach. Since farmers are already dealing with a lot of issues in the agriculture industry, our goal should be to make things easier for them. Applying innovative methods to farming may alleviate these issues. Machine learning techniques may be used to the agricultural sector. Crops may be analyzed using our clustering and classification algorithms.

Additionally, we may enhance agricultural yields by using regression approaches. So far, only the Naive Bayes and K-Nearest Neighbor methods have been taken into account in this research. These two approaches allow us to foretell which crops will thrive in a given environment and time of year. Because most farmers aren't familiar with the Weka tool, we built a Java app to help them out. They may anticipate the yield with the aid of this application. By entering the crop name, season, and location, this software allows us to conduct individual tests. Both the KNN and NB methods are at our disposal. You may choose the technique and start mining the results the moment you provide the input. You may learn the crop's yield rate from the findings. Dataset analysis also allows us to undertake multiple testing. You may choose a whole file at once and get accurate results when you use it for analysis. We may skip the steps of running individual tests and go straight to completing numerous tests here. Finding the precision between two approaches is made easier with this testing. This will help us choose which approach is the best out of the ones that are provided. Farmers will be able to use this information to better choose crops for their area. Results from prior years' data are included in the datasets. Predicting outcomes for new cases is much easier using these databases. Any given instance may be used by farmers to determine the crop output rate. Thus, this app aids farmers in choosing the best crop for their plot of land. Furthermore, it aids in the prediction of the crop's production rate. You may manually implement these methods. The values of the cases' probabilities are taken into account here. For new instances, we may get the outcome. To determine the likelihood of good and bad, the Naive Bayes approach is used. Plus, we can foretell whether the chosen crop will provide a high or low yield. In a similar vein, the KNN approach will measure the instances' distance from two provided values and then identify the least value. The distance between two values may be calculated using this approach by using the Euclidean distance. One. The Naive Bayes Method The naive bayes classifier is the foundation of the naive bayes algorithm. The likelihood of anticipated classes may be determined with the aid of this classifier. When creating massive databases, this approach is simple.



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

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.

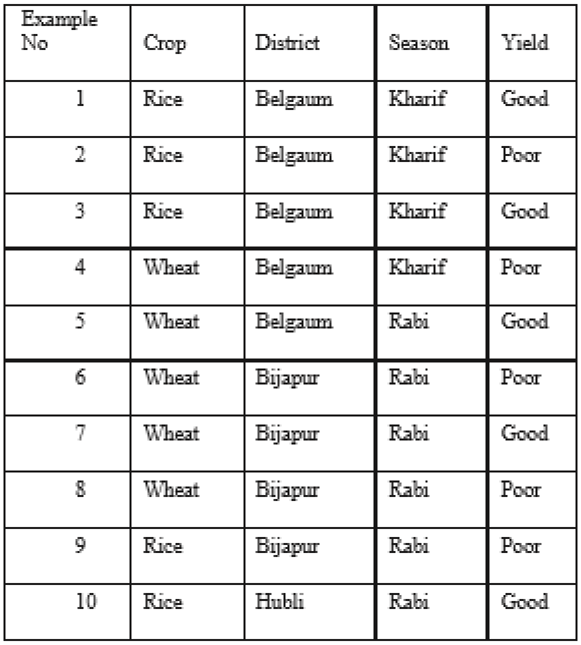
P(X) = probability of X

Naive Bayes Classifier

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

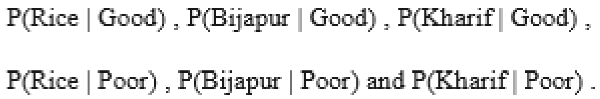
- n is number of training examples for which $v=v_j$.
- nc is number of examples for which $v=v_j$ and $a=a_i$.
- p is prior estimation for $P(a_i \mid v_j)$.
- m is equivalent sample size. Example:

TABLE 1: DATASETS FOR CROP YIELD PREDICTION



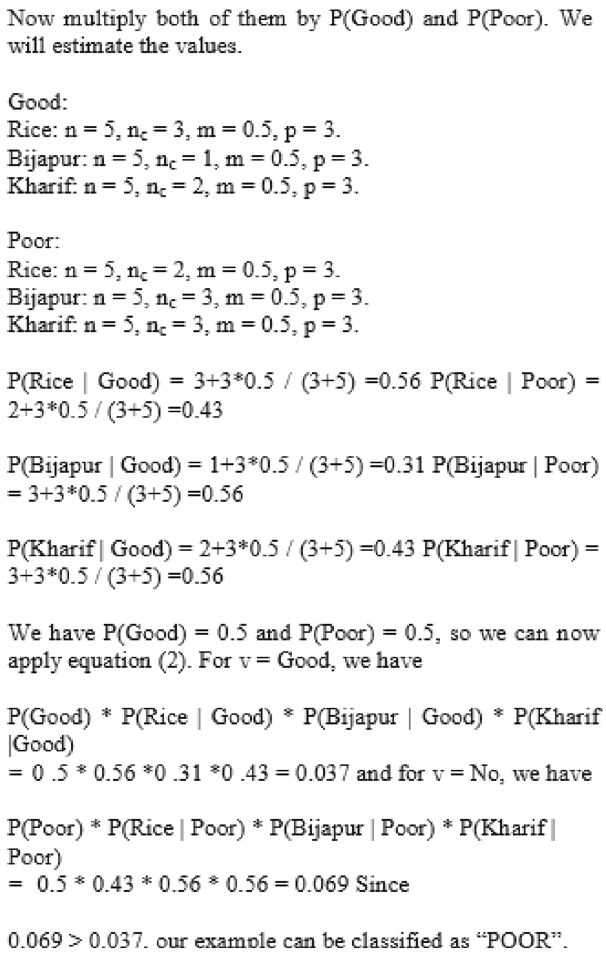
Example No	Crop	District	Season	Yield
1	Rice	Belgaum	Kharif	Good
2	Rice	Belgaum	Kharif	Poor
3	Rice	Belgaum	Kharif	Good
4	Wheat	Belgaum	Kharif	Poor
5	Wheat	Belgaum	Rabi	Good
6	Wheat	Bijapur	Rabi	Poor
7	Wheat	Bijapur	Rabi	Good
8	Wheat	Bijapur	Rabi	Poor
9	Rice	Bijapur	Rabi	Poor
10	Rice	Hubli	Rabi	Good

Training case study Kharif Bijapur Rice will be categorized. Rice Kharif Bijapur is not included in any of the datasets either. Now we can determine the likelihoods.



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. K-Nearest Neighbor (KNN) Analysis Both classification and regression predicting issues may be addressed with the K-nearest neighbor approach. This approach is useful for calculating time and predictive power, as well as for interpreting results. Machine learning methods have use in many domains. Another machine learning approach is KNN. Method of sample based learning is another name for this. You may use this to make predictions based on fresh datasets since it has data from previous datasets. Distance functions, such as the Manhattan or Euclidean distance, will be applied. You may use this to find the distance between your samples and every other training sample. It determines the desired outcome for fresh samples. A weighted total of the target values of the k closest neighbors will be used to determine the target value. The forecast may have a direct proportionate effect on the K-valve. When K's valve is tiny, it means that variation is large and bias is low. Low variance and strong bias are indicated by a K-valve that is greater than this. The lack of a need for training or tuning is the key benefit of this KNN. The fresh datasets are predicted by this KNN using data samples. This results in increased complexity and a longer processing time.

IV. SYSTEM ARCHITECTURE

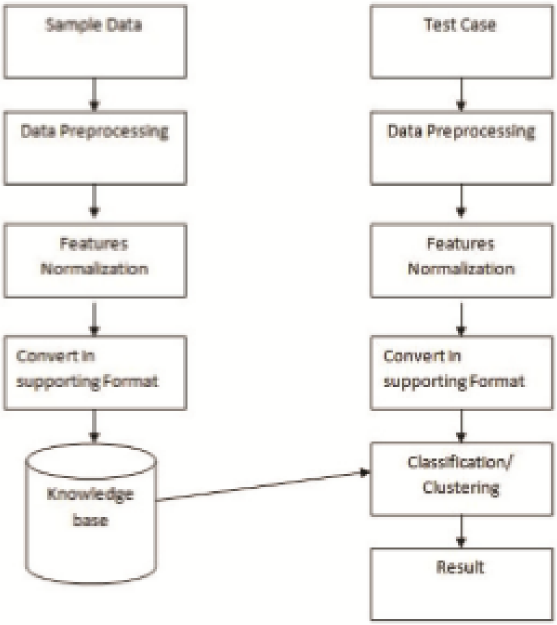


Fig. 1: System Architecture for Crop Yield Prediction

The initial step is to gather the data sets, after which we analyze them and eliminate any contaminants. The data is then transformed into a more manageable format, if necessary, by standardizing it. After that, the data is transformed into a format that may be used. Afterwards, it is deposited into the databases. The necessary procedure is then executed. The end findings are now in.

V. RESULTS

Predicting the rate of agricultural production requires the development of a Java program. All three components make up this application. Dataset management, testing, and analysis come in that order of importance. Datasets from prior years may be retrieved and transformed into a supported format via the management of datasets. Since this project makes use of the Weka tool, we have transformed all of the datasets into attribute relation file format. We may do individual tests in the testing phase. Two machine learning approaches have been investigated. The K-Nearest Neighbor approach and the Naive Bayes method are two examples. During testing, we have the freedom to choose any technique and conduct dataset analyses. For example, we may get yield results by picking a certain crop, location, and season. In the analysis section, we may compare the two approaches' accuracy by inputting a whole dataset file. This is useful for determining the best approach.

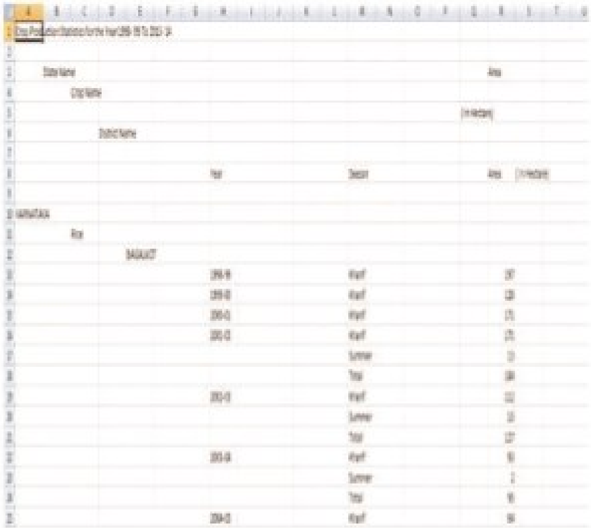
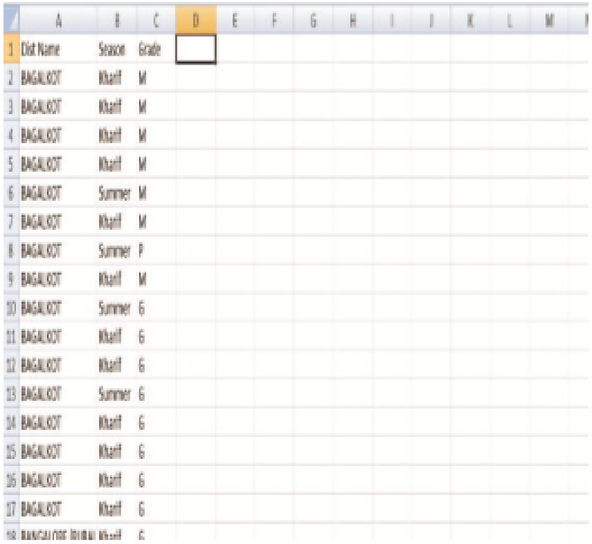


Fig. 2: Original datasets



	A	B	C	D
1	Dist Name	Season	Grade	
2	BAGALKOT	Kharif	M	
3	BAGALKOT	Kharif	M	
4	BAGALKOT	Kharif	M	
5	BAGALKOT	Kharif	M	
6	BAGALKOT	Summer	M	
7	BAGALKOT	Kharif	M	
8	BAGALKOT	Summer	P	
9	BAGALKOT	Kharif	M	
10	BAGALKOT	Summer	G	
11	BAGALKOT	Kharif	G	
12	BAGALKOT	Kharif	G	
13	BAGALKOT	Summer	G	
14	BAGALKOT	Kharif	G	
15	BAGALKOT	Kharif	G	
16	BAGALKOT	Kharif	G	
17	BAGALKOT	Kharif	G	
18	BANGALORE RURAL	Kharif	G	

Fig. 3: New datasets

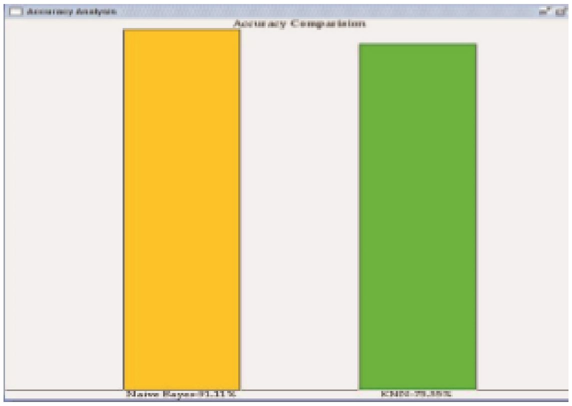


Fig. 4: Result of checking accuracy

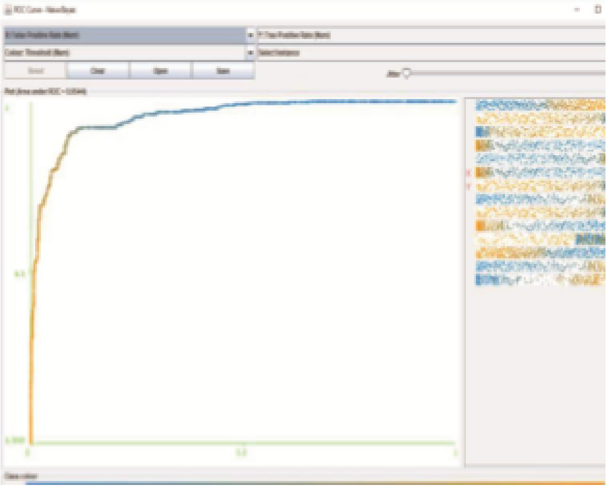


Fig. 5: ROC curve for the results

VI. CONCLUSION

Our country's economic prosperity is aided by the agricultural sector. However, when it comes to using emerging machine learning technology, this falls short. Therefore, it is imperative that our farmers be well-versed in the latest machine learning and other cutting-edge methods. To get the most out of your harvest, use these methods. In order to increase agricultural yields, several machine learning methods are used in agriculture. Additionally, these methods are useful for addressing agricultural issues. By comparing several approaches, we may also determine the yield's accuracy. Because of this, we may enhance performance by comparing crop yields to ensure accuracy. A wide variety of agricultural applications make use of sensor technology. If you want your crops to produce as much as possible, this document is for you. Additionally, it assists them in choosing the best crop to grow on their chosen property during the chosen season. When used to the agricultural sector, these methods will alleviate farmers' hardships. Our country's economic development will be boosted by this.

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