Crop Prediction using Naive Bayes and Decision Tree Machine Learning Algorithms

[1] Mallegowda M [2] Dr. Anita Kanavalli [3] Manish V, [4] Mohammad Rahil KA, [5] Achyuth Kumar K
[1] Asst. Prof. Department Computer Engineering
[2] Head of Computer Science Department
[3,4,5] UG students, Department of Computer Science and Engineering
MS Ramaiah Institute of Technology, Bangalore
Affiliated to VTU

Abstract: *In this country, as we know, agriculture is the backbone.Agriculture accounts for approximately 17% of India's GDP (gross domestic product (gdp) and provides employment to over 60% of the population, as it is the main occupation of Hindus, since the earlier days. Our paper is suitable for almost all crops in India. In this model, we made the effective use of different parameters, such as the state, the region, as temperature, pH, water, etc., etc. and each user will have to select the most suitable crops within a certain period of time. Agriculture also plays an important role in the economic sector of our country. In this sector of the economy, giving rise to serious concerns, and it is becoming more and more a topic that is relevant to everywhere. The usual and traditional methods that are being used by farmers are not enough to satisfy their needs and wants. Therefore, new methods to predict crop has been introduced. These methods will not only meet the needs of the farmers, but also to provide employment to millions of people. Machine learning has led to a revolution in agricultural, real-world applications. In this machine learning techniques, produced a variety of enhancements, such as employment and climate change.*

Keywords: *Crop prediction, Machine learning, Gaussian Naive bayes, Decision tree, KNN, Agriculture, temperature, pH, Rainfall, Humidity, Farming.*

1.Introduction

Agriculture plays a very important role in the lives of every individual. In our model, we make use of different parameters, such as temperature, pH, and information about its location, and so on. There are a lot of areas in India. The plants you'll be better seen and understood by those methods. With all of the relevant information by the Authorities. This dataset consists of features such as temperature, humidity, soil type, soil pH, rainfall, and location, etc.). I have also used machine learning techniques such as Naive Bayes, Decision Tree. Became one of the most important decisions that you will have to make in order to determine which crops will need to be placed in the appropriate plants.Growing crops has a very important role to play in the development of the agricultural sector. India has the third largest economic country in the world just behind the United States and China. And, with India in second place in the production of crops such as rice, wheat, sugar cane, peanuts, etc.).

2.Literature Review

Our teams have a lot of research on different types of paper, and many other sources on the Internet, and a review of the literature concerning them, we have to go through the following scenarios and methods of their solution and development, etc.

[1] In this system,they used SVM and Decision Tree algorithms to get the crop predictions.The objective is to use various types of machine learning algorithms to get the crop prediction.

[2] In another study proposed a machine-learning system that provides for the right of crop collections with the help of machine learning algorithms.

[3] Its system that provides for the right set of crop collections with the help of a supervised decision-tree algorithm that is a simple moving average algorithm.

[4] They have used K-Nearest-Neighbours, Multinomial Naive Bayes Algorithm and Decision Tree Classifying Algorithm, and it was observed that the decision tree classifying algorithm gives a better crop of predictions for their collection. The improvement of this project will involve the collection of data by means of the feedback of the users, and the use of such data for future predictions.

[5] In this paper,they have made a comparison between many algorithms. They used many algorithms such as KNN, KNN-cross, SVM and many other algorithms.

[6] Precision agriculture is a state-of-the-art technology for agriculture, which uses data from the ground to study, and provides farmers with the right crop on the basis of the parameters you specified.They developed a recommendation system based on the machine learning algorithms, such as Random Tree, CHAID etc. in order to have a crop of recommendations.They compared Naive Bayes, Random Tree, KNN, and CHAID. The use of a resembling method, which combines the effect of the use of two or more different models to better predict the crop yield.

[7] In this paper,the author suggests that the production of the products are tested and rated in accordance with the study of the rate of return. This rank is calculated using various data mining algorithms. They used a diverse group of principles, such as K-nearest Neighbors(KNN), and Classifying Algorithm Naive Bayes . With the help of this idea, they made an estimate of the classification rules, and in particular, in accordance with the data set they used accordingly in projects.

[8] In this paper the author uses a data set that was pre-processed using the basic pre-processing tasks. The recommendations for the cultivation of the plants used in naive Bayes, and J48 classifiers. The final recommendations are made with the help of association policy based on the outcome of the classifiers.10 Cross Validation was used to train the model. The tests were carried out on the basis of various indicators, such as Accuracy, ROC area, Listing, Accuracy, F-measure etc.

[9] The experiment was conducted by combining various Machine Learning Algorithms and using that combination to obtain the prediction. The future work in this project will be to apply various algorithms for the dataset and to improve the accuracy even more.

[10] The dataset is prepocessed and then it breaken down into a smaller subset and simultaneously an associated decision tree is developed.This classifying decision tree handles both the numerical

and categorical data. ID3 is the main technique that is being used here.It uses a greedy top-down search approach and there is no backtracking involved.That's the reason for this technique to be very fast.

[11] The dataset is pre-processed so that they apply a technique produces multiple decision trees during training phase of the algorithmRandom forest algorithm is used to get better results.They create decision trees on data sets and then gets the prediction from each one of the trees.Finally,a the optimal/best solution is chosen by voting algorithm.After this project uses an ensemble technique. This technique is better than decision tree because it reduces the complexity of the program.

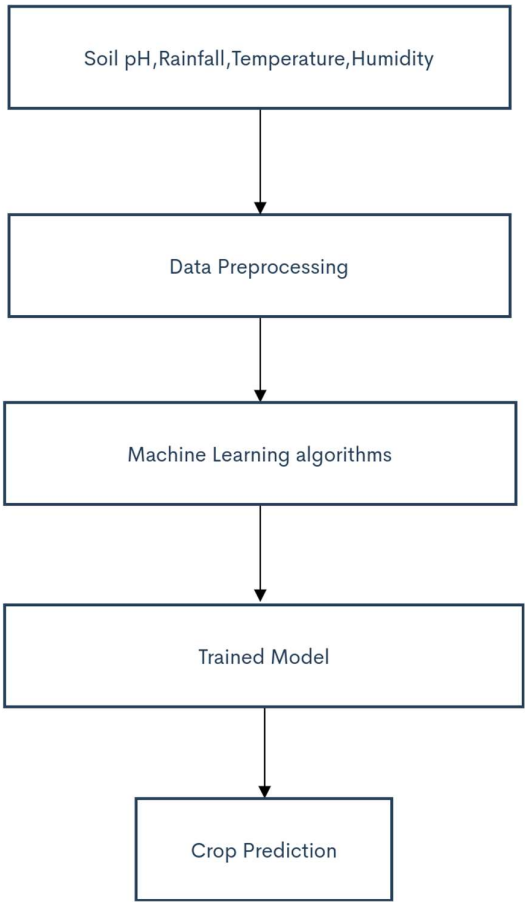
[12] A team of researchers used clustering as the main technique.The main objective is to find a partition an data records into clusters such that points within each cluster are nearer to one another. A process involving machine learning may involve many steps and may not be linear to obtain efficient results.In later stages , each model implementation only accepts data in its own individual format, and has its own ways of specifying parameters and gives and efficient output.

[13] Initially all the research queries are well defined. When research queries are defined databases will be used to select the appropriate studies.After the selection of the relevant studies they are filtered and extracted.The approach can be followed in 3 parts those are plan conduct and report review to obtain appopriate results.
In future stages the information retrieved will aim to determine whether or not the extracted studies meet the standard of exclusion criteria and produce an accurate results.

[14] In a paper another research suggested that we have to decompose the training data with the help of class values so that attribute/feature value will map to the suitable data.Summarized data tuples can be obtained by calculating mean and Standard deviation for needed tuple.Probabilities can be obtained by comparing summarized data list and original list. The result with higher probability will be taken for crop prediction.
Random-Forest-Decision algorithm is best crop prediction with low number of dataset.This technique is suitable for crop prediction in agricultural fields which will create an impact. In future they can determine the appropriate algorithm which can be used for prediction will be based on accuracy.It describes the crop prediction ability of an algorithm.

[15] In this paper they have used good agriculture data-set for crop prediction.The regression and classification analysis is done for the data-set.Basically Data Analysis is done through regression analysis.The implementation of these models in future stages will help in better cultivation in agricultural field,further it can reduce the loss faced by farmers by predicting correct crops.

3.Proposed System/Methodology



The project proposes a machine learning model which can do the crop predictions based on various factors such as temperature,soil pH,humidity and rainfall.

A. Training dataset:

We need to have a good dataset in order to get good results.Our dataset contains temperature,rainfall,humidity and pH values. Thus, by using a correct machine learning algorithm which suits the best we can train the dataset and get the crop predictions that can be grown under the given circumstances.

B. Data Pre-processing

Data set should be preprocessed before hand so that we can apply algorithms on them and get the desired results.A dataset might have missing values in them,we have to handle the missing values in this step.Also we have to do necessary feature scaling and normalization if required.

C. Machine Learning Algorithms

We have used two machine learning algorithms in this project.The first one is Gaussian Naive Bayes Machine Learning Algorithm and the second one is Decision Tree Algorithm.We trained our Machine learning algorithms using both these algorithms and found out that Naive Bayes Gives Better results.So in our web application backend we use Naive Bayes Algorithm to give the results to the user.

1.Gaussian Naive Bayes Classifier

Working on Naive Bayes classifier:

1.Convert the given data set into frequency tables i.e Represent the data in Normal Distribution curve.

2.Generate a Probabilistic-Likelihood table/matrix by computing the necessary probabilities of provided features.

3.Apply Bayes algorithm to compute the output(Posterior Probability).

Working on Gaussian Naive-Bayes classifier:

1.For all the features we calculate the mean and standard deviation and draw the corresponding Gaussian Distribution curve for all the features with respect to the crop names.

2. If some new data is given to the model, the model performs the following activities.

First Prior Probabilities (Initial Guess) is found out for all the crops.Lets say P(A).

Then for finding the posterior-probabilities we use the formula,

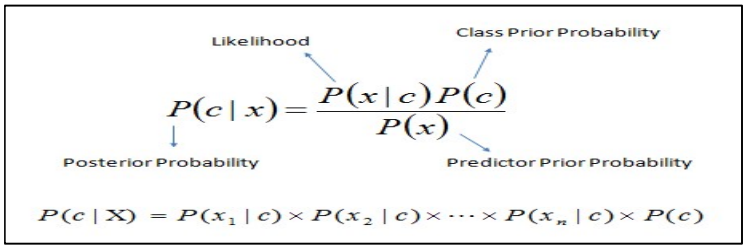
```
P(A|B) = [P(B|A) * P(A)] / P(B)
```

In our case the formula can be seen as,

P(maize | rainfall,pH,temperature,humidity) = [P(maize) * P(rainfall = 200 | maize)

* P(pH = 7|maize) * P(temperature = 25|maize) * P(humidity = 85|maize)]

Now compare all the Posterior Probabilities for all the crops and predict the crop which has higher probability values.



2.Decision Tree Classifier

Working of Decision tree classifier:

1.Initially the root node has the complete dataset to classify further. We have to keep splitting the dataset recursively to maximize the Information Gain.

Entropy is the measure of Information contained in the state.

Entropy = $\sum$ -p(i) [log(pi)]

p(i) => probability of class i.

Calculate probabilities for all possible states pick the lesser entropy state.i.e, Information Gain = Entropy of Parent - Summation of Entropies of Child.

Whichever information gain is higher pick that and proceed further.

Finally at some time we get Entropy 0 which means its a pure class.

i.e we have found the final crop name class.

One thing to note down is that,its a greedy algorithm and there is no backtracking involved here.

D.Trained Model

Trained Models are obtained when we apply these machine learning algorithms to the dataset.Here we applied two algorithms,Gaussian Naive Bayes and Decision Tree Machine Learning Algorithms and obtained two trained models.We also tested both these models with 30% splitted data set and obtained accuracy,F1-Score for both the algorithms.

4.Result And Discussion

From the literature survey and various research from many other sources ,we have found out that,Prediction of a crop mainly depends on the rainfall,temperature,soil pH and humidity.

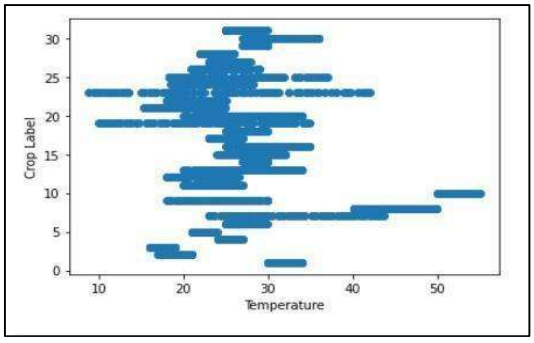


Figure 1. Scatter graph between Crops and Temperature

These factors were found to be the necessary which should be taken into consideration than

some other factors like type of soil or type of crops grown before which was found to be of least importance.

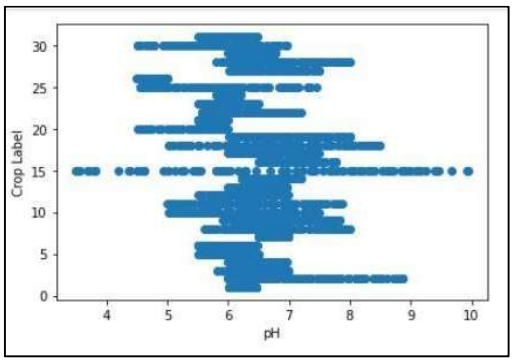


Figure 2. Scatter graph between Crops and pH

After our survey we understood that crop prediction using machine learning is still a challenge in this field and there are many experiments that are been made in crop prediction using machine learning.From our research and work on the project we have found that 3 algorithms were best suited for crop prediction namely KNN,Decision Tree Algorithm and Naive Bayes algorithm.But we must be sure that we must pick the best one which gives us the most efficient output on which our users can be reliable on, thus by comparing these algorithms we found out Naive bayes to be the most efficient which gives the best output on which user can rely on.Thus we chose Naive bayes algorithm for our project.

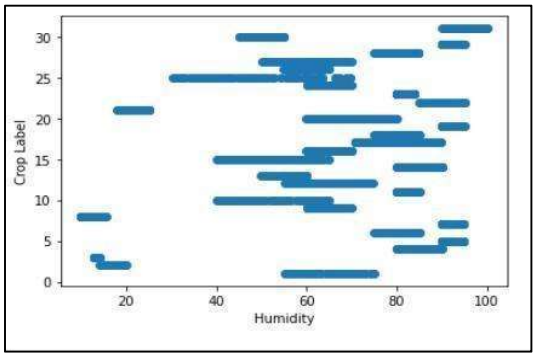


Figure 3. Scatter graph between Crops and Humidity

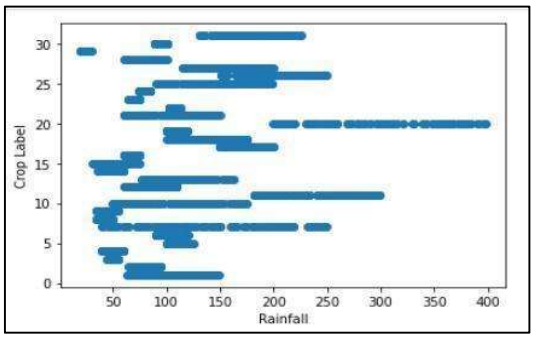


Figure 4. Scatter graph between Crops and Rainfall

Scatter graph's for the different crops has been shown.Basically The scatter diagram graphs

are graph between numerical data.We will have 2 variables maked in X and Y axis.Through this graph we can understand the relationship of variables and the data.

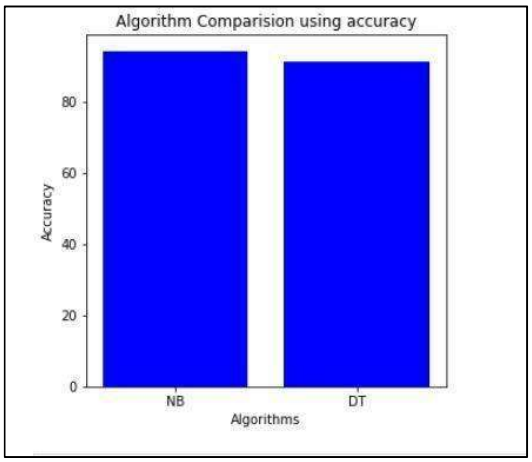


Figure 5. Algorithm Comparison using Accuracy

In the Figure 5 we can see that Accuracy of Naive bayes is better than that of Decision tree,Thus we choose Naive bayes over Decision Tree.We also got good F1-score for Naive Bayes Algorithm comparatively.So Naive Bayes works perfect in our case.

5.Conclusion and Future scope

Farming is evolving as a profession nowadays and there are many works that are going related to Agriculture and Technology.Since agriculture and technology is evolving day by day,there are many systems that are made in order to improve the farming experience.The proposed system recommends the best appropriate crop by considering various parameters as temperature,humidity,rainfall and soil pH.Our crop prediction project will help the farmers to take decisions wisely and estimate the crop based on weather parameters,using which farmers can grow the appropriate crop.These methods shows the scope of research and availability of data.These methods also differ in scale,area,particular district ,crop etc.These methods also stated that models with great features did not provide best prediction of a particular crop.Future work of this kind may be a well developed system which can include some more features where user can give some inputs based on facilities he has in his farm(like irrigation) to grow a crop.So that he can get better results.

References

[1]Mahendra N,"Crop Prediction using Machine Learning Approaches" 2020,International Journal of Engineering and Technical Research.

[2]Mythresh,Lavanya,Meghana,Nisarga,"Crop Prediction using Machine Learning",2020,International Research Journal of Engineering and Technology (IRJET).

[3]"Crop Prediction System using Machine Learning", International Journal of Emerging Technologies

and Innovative Research (www.jetir.org | UGC and issn Approved), ISSN:2349-5162, Vol.6, Issue 5, page no. pp 178-182, May 2019.

[4]Pavan Patil, Virendra Panpatil, Prof. Shrikant Kokate,"Crop Prediction System using Machine Learning Algorithms",2020.

[5]Kevin Tom Thomas,Varsha S,Merin Mary Saji,Lisha Varghese,Er. Jinu Thomas,"Crop Prediction Using Machine Learning",2020,International Journal of Future Generation Communication and Networking.

[6]S.Pudumalar,E.Ramanujam,"Crop Recommendation System for Precision Agriculture",2016 IEEE Eighth International Conference on Advanced Computing (ICoAC).

[7]R.Kumar,M.P.Singh,P.Kumar and J. P. Singh,"Crop Selection Method to maximize crop yield rate using machine learning technique",2015.

[8]Viviliya B,Vaidehi V,"The Design of Hybrid Crop Recommendation System using Machine Learning Algorithms",2019.

[9]T.R.Lekhaa,"Efficient Crop Yield and Pesticide Prediction for Improving Agricultural Economy using Data Mining Techniques" ,2016.

[10]Akshay Prasanna S,B A Harshanand,B Srishti,Chaitanya R,KirubakaranNithya Soundari,SwathiSriram,V Manoj Kumar,VarshithaChennamsetti,Venkateshwaran G,Dr.Pramod Kumar Maurya VIT University,Vellore,Tamil Nadu,India.

[11]International Journal of Future Generation Communication and Networking Vol.13,No.3s,(2020),pp.769–772.

[12]Review of Literature of data mining techniques for crop yield prediction.

[13]Crop yield prediction using machine learning: A systematic literature review Author links open overlay panelThomasvan KlompenburgaAyalewKassahunaCagatayCatalb.

[14]P.Priya,U.Muthaiah & M.Balamurugan, International journal of engineering sciences & Research Technology predicting yield of the crop using Machine Learning Algorithm.

[15]Sangeeta, Shruthi G International journal of scientific & Technology research volume 8,Issue 01,January 2020 ISSN 2277-8616 544 IJSTR©2020 www.ijstr.org Design And Implementation Of Crop Yield Prediction Model In Agriculture.