WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Naive Bayes and Decision Tree Machine Learning Algorithms for Crop Prediction

[1]S. Revanth Babu,[2]L.Charitha

[1,2]Asst. Prof. Department Computer Science Engineering,Annamacharya Institute of Technology and Sciences(Autonomous),Tirupati.Andhra Pradesh

ABSTRACT:

We all know that agriculture is the foundation of this nation.Over 60% of India's population is employed in agriculture, which contributes roughly 17% of the country's GDP and has historically been the primary vocation of Hindus. Almost all Indian crops can be used with our paper. We effectively used a variety of parameters in this model, including state, region, temperature, pH, water, etc. Each user will need to choose the best crops within a specific time frame. A significant part of our nation's economy also involves agriculture. This area of the economy is raising significant issues, and it is increasingly a topic that is pertinent to everyone. Farmers' demands and wants cannot be addressed by the customary and conventional ways they currently employ. hence, fresh approaches to Predict crop has been made available. These techniques will not only fulfill the requirements of the farmers, but to provide jobs for millions of people. Agriculture has undergone a transformation because to machine learning applications in the actual world. The use of machine learning techniques in this study led to a number of improvements, including as well as climate change and jobs.

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

1. Introduction

Every person's life is profoundly impacted by agriculture. We incorporate a variety of elements into our model, such as temperature, pH, geographical data, and so forth. India has many different regions. These techniques will enable you to view and comprehend the plants better. In light of all the the Authorities providing pertinent information. This dataset includes characteristics like temperature, humidity, the location, rainfall, soil type, soil pH, etc.). Additionally, I have employed machine learning strategies such Decision tree and Naive Bayes.become one of the most crucial choices you will need to make in order to choose which crops should be planted in the right plants.The development of the agricultural industry depends heavily on the production of crops. The third largest country is India. The third-largest economy in the world after China and the United States. India came in second place. crops including rice, wheat, sugar cane, peanuts, and other crops are produced.

2. Literature review

Our teams have done extensive study on many forms of paper, as well as other sources on the Internet and a review of the literature on them. We must also go over the following situations and the approaches to solving and developing them, among other things.

[1] To obtain crop predictions in this system, SVM and Decision Tree algorithms were applied.The goal is to anticipate the crop using a variety of machine learning algorithms.

[2] In a different study, a machine-learning system was put out that, using machine learning algorithms, provides for the right of crop collections.

[3] Its system uses a supervised decision tree method, which is a simple moving average technique, to give for the appropriate set of crop collections.

[4] It was shown that the decision tree classification technique produces a superior crop of predictions for their collection when compared to K-Nearest Neighbors, Multinomial Naive Bayes, and other algorithms. This project will be improved by gathering data from user feedback and using that information to make predictions about the future.

[5] In this study, they compared a variety of algorithms. They employed numerous algorithms, including KNN, KNN-cross, SVM, and many more.

[6] Modern agriculture uses precision agriculture, a cutting-edge technology that analyzes data from the ground and gives farmers the perfect crop based on the parameters you selected.

[7] In this essay, the author proposes that product creation be evaluated and assessed in accordance with a research of return rates. Several data mining algorithms are used to calculate this rank. They applied a wide range of theories, including K-nearest Neighbors (KNN) and the Naive Bayes Classifying

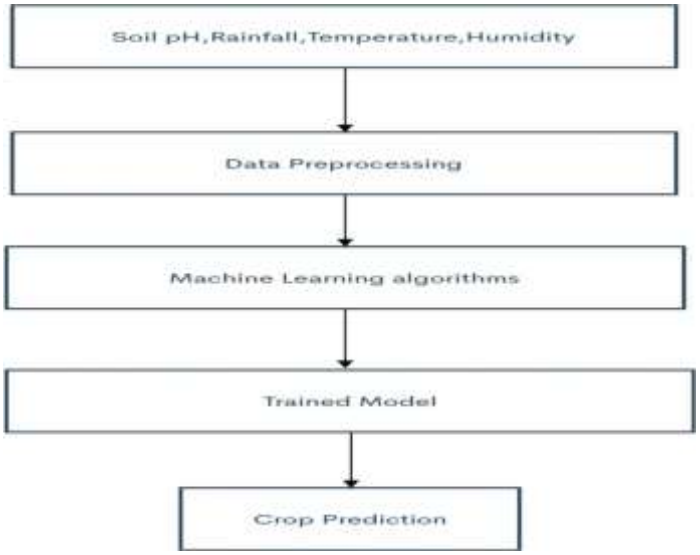
Algorithm. With the aid of this concept, they estimated the categorization criteria, specifically in light of the data set they employed appropriately in projects.

[8] The pre-processed data set used in this paper was created utilizing the fundamental pre-processing activities. the suggestions for growing the plants that were employed in J48 and naive Bayes classifiers.

[9] Combining different machine learning algorithms was done in the experiment in order to get the prediction. Future work on this project will involve applying multiple algorithms to the dataset in order to further increase accuracy.

[10] The dataset is prepared, then broken down into a smaller subset while simultaneously developing an associated decision tree. Both numerical and category data are handled by this categorizing decision tree. The primary strategy employed in this case is ID3.There is no backtracking involved and it uses a greedy top-down search method. This is why this method is so effective.

3. Proposed System/Methodology



The study suggests a machine learning model that can anticipate crops based on a number of variables, including temperature, soil pH, humidity, and rainfall.

A. Training dataset:

In order to acquire decent results, we need a solid dataset. Temperature, rainfall, humidity, and pH variables are all included in our dataset. Therefore, by training the dataset with the best machine learning algorithm, we may obtain crop forecasts that can be cultivated under the given conditions.

B. Data Pre-processing:

Data sets need to be preprocessed in order to apply algorithms to them and obtain the intended outcomes.In this phase, we must deal with any missing values that may be present in a dataset. Additionally, if necessary, we must scale and normalize the features.

C. Machine Learning Algorithms:

In this project, we used two machine learning techniques. The first one is the Decision Tree Algorithm, while the second is the Gaussian Naive Bayes Machine Learning Algorithm.We trained both of these algorithms in our machine learning algorithms and discovered that Naive Bayes produces superior results.In order to provide the user with the findings, we employ the Naive Bayes algorithm in our web application backend.

1.Gaussian Naive Bayes Classifier:

Working on the Naive Bayes classification

1. Convert the provided data set into frequency tables, or present the data as a normal distribution curve.

2. Create a probabilistic-likelihood table or matrix by calculating the required probabilities for the features that are given.

3. Use the Bayes algorithm to calculate the result (Posterior Probability).

Working with the Gaussian Naive-Bayes classification:

1.We determine the mean and standard deviation for each feature and then create the matching Gaussian Distribution curve for each feature in relation to the crop names.

2. If the model receives some fresh data, it carries out the following tasks.

For each crop, the First Prior Probabilities (Initial Guess) are determined.Suppose P(A).

The posterior probabilities are then determined using the following formula:

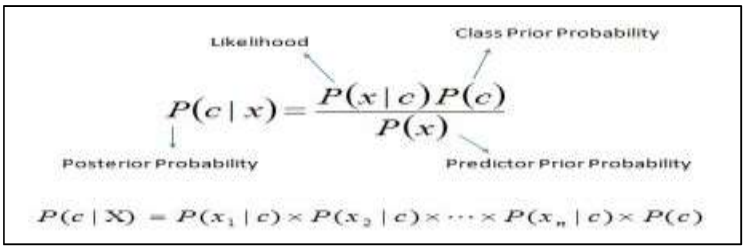
P(A|B) = [P(B|A) * P(A)] / P(B).

P(maize | rainfall, pH, temperature, and humidity) can be interpreted as the following in our case:

[P(maize) * P(rainfall = 200 | maize) *

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

Predict the crop with the highest posterior probabilities by comparing all the posterior probabilities for all the crops.



2. Decision Tree Classifier:

Workings of the Decision Tree Classifier: 1. The root node initially contains the entire dataset for further classification. To maximize the Information Gain, we must continue recursively partitioning the dataset.

The amount of information present in a state is measured by its entropy.

Entropy is equal to -p(i) [log(pi)].

p(i) = likelihood of class i.

Pick the state with the lowest entropy after computing the probabilities for each potential state.For example, Information Gain =

Parental Entropy plus Child Entropies Added.

Choose whichever information gain is greater and carry on from there.

Finally, at some point, we reach Entropy 0, which indicates that the class is pure.

That is, we have identified the last crop name class.

One thing to keep in mind is that this algorithm is greedy and that there is no backtracking involved.

D. Trained Model:

When we use these machine learning algorithms on the dataset, we get trained models.Here, we used the Gaussian Naive Bayes and Decision Tree Machine Learning Algorithms to apply two algorithms and produce two trained models.With a 30% split of the data set, we also evaluated both of these models and obtained F1-Score for both algorithms, accuracy

4. Result and Discussion

We have discovered that the selection of a crop primarily depends on the amount of rain, the temperature, the soil's pH, and the amount of humidity after doing a literature review and conducting several other types of research**.**

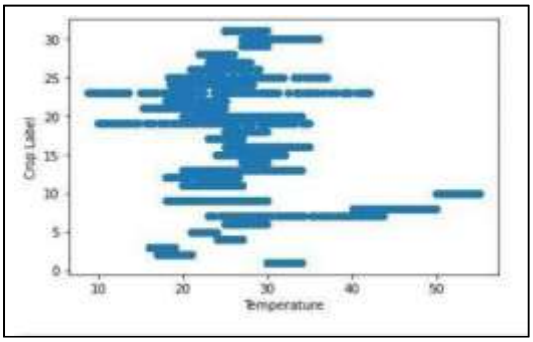


Figure 1. Scatter graph between Crops and Temperature

These factors were discovered to be necessary and should be taken into consideration rather than some other factors, such as the kind of soil or the type of crops that were grown previously, which were determined to be of minimal significance.

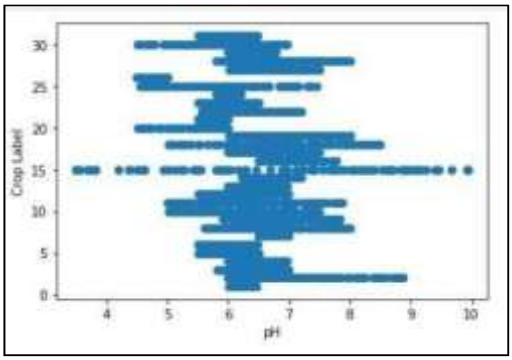


Figure 2. Scatter graph between Crops and pH

After doing our survey, we came to the conclusion that crop prediction using machine learning is still a challenge in this sector and that many experiments have been made in crop prediction using machine learning.From our research and work on the project, we have determined that three algorithms—KNN, Decision Tree Algorithm, and Naive Bayes Algorithm—were best suited for crop prediction.By comparing these algorithms, we determined that Naive Bayes is the most efficient and provides the best result that users can depend on. However, we must be certain that we must choose the best option that gives us the most efficient output that our customers can be dependable on.Therefore, for our project, we settled on the Naive Bayes algorithm.

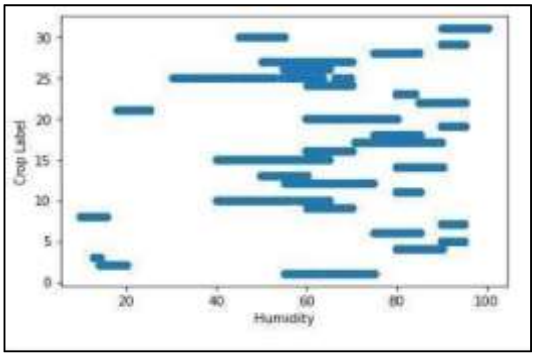


Figure 3. Scatter graph between Crops and Humidity

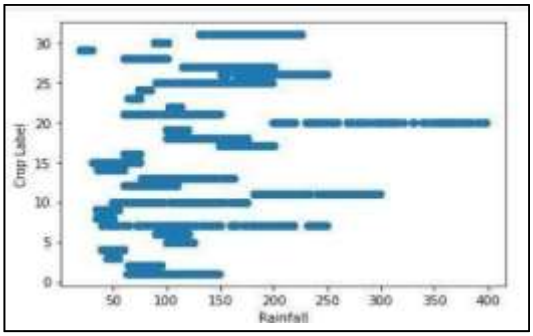


Figure 4. Scatter graph between Crops and Rainfall

The scatter graphs for the various populations have been displayed.Basically The spreadsheet diagram graphs are plotted between numerical data. Two variables will be made into X and Y axes.We may comprehend the relationship between the variables and the data using this graph.

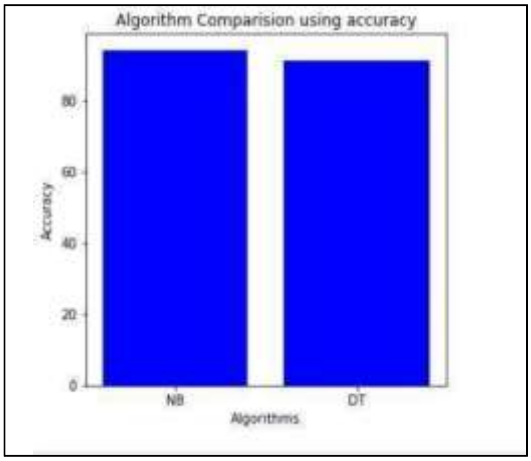


Figure 5. Algorithm Comparison using Accuracy

Because the accuracy of Naive Bayes is higher than that of Decision Tree, as shown in Figure 5, we prefer Naive Bayes to Decision Tree. Additionally, we obtained a respectable F1-score for the Naive Bayes algorithm. So, in our situation, Naive Bayes works perfectly.

5.Conclusion and Future scope

Today, farming is evolving as a profession, and many projects relating to agriculture and technology are in the works. Since agriculture and technology are developing day by day, several methods are being developed to enhance the farming experience. The proposed approach recommends the most appropriate crop by taking into account a number of parameters, including temperature, humidity, precipitation, and soil pH.Our crop production project will assist farmers in Make intelligent decisions and estimate the crop based on weather parameters, with the help of farmers can grow the appropriate crop. These techniques demonstrate the scope of research and the accessibility of data.These methods also vary in scale, area, particular distribution, crop, and other factors. They were also established. Those models with great features did not offer the best predation of a specific crop. Future work of this kind may consist of a fully developed system with additional capabilities that allow the user to provide input based on the farm's resources (such irrigation) to grow a crop.so he can acquire better outcomes.

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 Technologiesand 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.