

A Review of Machine Learning Algorithm used in Crop Recommendation

Ranjeet Sonkar¹, Yash Tyagi², Aditi Arora³ and Anju Joshi⁴

¹⁻⁴ABES Engineering College, Ghaziabad, India

Email: Ranjeet.20b0101065@abes.ac.in, yash.20b0101191@abes.ac.in, aditi.arora@abes.ac.in, anju.joshi@abes.ac.in

Abstract— In India, Agriculture plays a pivotal role in the Indian economy, contributing approximately 17% to the total GDP and employing over 60% of the population. Not only in India, agriculture sector is one of the most important sectors in all over the world. In the past few years many machine learning algorithms are being used to recommend the best crop for farmers under the different conditions. It is important to recommend the best crop to the farmers not only for the better growth but also for the decrease the risk of farmers. In this paper we are doing the review of different machine learning algorithms like Logistic Regression, Naïve Bayes, Support Vector Machine, Extra Trees, Decision Tree, Random Forest and many more. And then we find out the best algorithms based on their accuracy rate and used that algorithm in our project to recommend the crop. It also requires the dataset to train our system. So, we take the dataset of values like N, P, K, temperature, humidity, ph, and rainfall.

Index Terms— N, P, K, temperature, humidity, ph, rainfall; Machine Learning; Crop recommendation; Logistic Regression; Naïve Bayes, Support Vector Machine, K-Nearest Neighbors, Decision Tree, Random Forest, Bagging, AdaBoost, Gradient Boosting, Extra Trees.

I. INTRODUCTION

As the global population continues to expand, the demand for food production intensifies, necessitating more efficient and sustainable farming practices. So, in this context agriculture becomes one of the most important sectors. In India, as per the report, agriculture plays a pivotal role in the Indian economy, contributing approximately 17% to the total GDP and employing over 60% of the population. So, we can imagine that the impact of the agriculture sector is very large. Traditional farming practices rely on generational knowledge and local experience. But the current scenario has changed, our climate changes a lot and the soil characteristics also change a lot. By using the traditional method, it would be difficult to choose the best crop for cultivation, and this also enhances the risk for farmers. As we all know, machine learning has developed a lot and right now, we have many machine learning algorithms. And to cultivate a good crop which has low risk from the farmers' point of view we are using machine learning algorithms. Because the traditional methods are not very successful at today's time and in today's time everything is changed like climate changed, weather conditions changed, and soil characteristics have been changed. So, we have to be developed as science and the new technologies are being developed or introduced. So, in the current scenario, we have used the machine learning algorithms that will help the farmers to choose the right crop for their cultivation. Now, another important task is to choose the best machine learning algorithm and make the model that uses that algorithm which will give the more accurate result. Also, we want the data to train our system and use that data to see the result. We can take the data online or we can take the

real-time data as well. This paper contains a description of about crop recommendation and the different machine learning algorithms that can we can use for recommendation of crop. In this we have created a model which first check the accuracy of different algorithms and find out the best algorithm for recommendation of crop under the different conditions like climate, soil characteristics and ph.

A. Basic Concepts

What is Machine Learning?

Machine Learning Algorithms are mathematical models that have various techniques and formulas that are used on certain parameters or raw data to provide meaningful results and pieces of information. Machine Learning Algorithms are the foundational block for data science, artificial intelligence, and other technical fields. There are many applications of machine learning algorithms which are self-driving cars, recommendation techniques, recognition of images, prediction of diseases, etc. The end goal of machine learning algorithms is to develop a model that gathers information and data automatically and uses it for training the model to get better accuracy in predicting the result. In machine learning algorithms, there are four types of algorithms which are Supervised Learning Algorithms, Unsupervised Learning Algorithms, Semi-supervised Learning Algorithms, and Reinforcement Learning Algorithms.

What is Crop Recommendation?

A crop recommendation is a model that refers to advice or guidance provided to farmers for selecting the suitable crop for cultivation under different conditions like environmental, climate and soil. This recommendation is typically used the machine learning algorithms to recommend the crop. There are many machine learning algorithms like Random Forest, Logistic Regression, Naïve Bayes, Decision Tree and many more. Key factors considered in crop recommendation include: - Climate and Weather Conditions, Soil Type and Quality, Water Availability, Pest and Disease Resistance, Market Demand, Crop Rotation and Diversification, Technology and Farming Practices. The goal is to help farmers optimize their yields, reduce risks, and enhance the overall sustainability of agriculture practices in a given areas.

B. Methodology

To build a crop recommendation model, we have to follow some steps to make it possible and get accurate outcomes. The task is broken into seven steps which are shown below:

Explore Dataset

The dataset is the key by which the machine learns and train himself to give the next result. Data must be reliable so that machines can produce accurate results by acknowledging the patterns correctly.

Data Preparation

After we get our reliable dataset, we have to perform some actions which are: Gather all the data and check whether it would be sufficient for our model or not and the data must be checked before the implementation that it would not contain any null values, there is no missing rows and columns.

Model Choosing

It is the most important thing to select an appropriate algorithm and model by which we can get the most accurate result. We have taken many algorithms in this model and then choose one of them which has the highest accuracy rate.

Model Training

Training of machine would be the most important part in the creation of any project. So, in this we have taken many machine learning algorithms and check their accuracy rate and then we choose the best one to train our model and that will be used in the model for giving the accurate result

Model Prediction

After training, and evaluating the best accurate rate algorithm, we have to train our model over the dataset and then see the result of it.

C. Literature Review

S. Kanaga S. R., et.al. [1] has proposed a paper in which they are using non-linear regression technique to predict crop yield and price that a farmer can obtain from his land. In this, they are predicting based on different factors that are affecting agricultural production like rainfall, temperature, market prices, area of land and past yield of a

crop. Here, this model involves many steps, firstly collecting the dataset that includes crop areas, types, nature of soil and many more. After that, crop demand prediction is classified by changing the crops' market prices. After that suggestion of crops is done on data mining results, the various parameters that are considered for the crop recommendation are total cultivation of crops, area, overall consumption, and the demand grade. In this model a text translation feature was also being introduced to convert text from English to regional language. This system overall reduces the loss faced by the farmers today and would improve the yield for the farmers.

Priyadharshini A., et.al. [2] has proposed a paper in which they have proposed a system to assist the farmers in crop selection by considering all the factors like soil, geographical location, and sowing season. The main objective of this system is building a robust model that gives a correct and accurate prediction of crop and that provides profit analysis of various crops based on the previous year's data. The system is being implementing by using machine learning algorithms like Decision Tree with 81% accuracy, K Nearest Neighbor with 85% accuracy, K Nearest Neighbor with cross validation with 88% accuracy, Linear Regression Model with 88.26% accuracy, Naïve Bayes with 82% accuracy, Neural Network with 89.88% accuracy and Support Vector Machine with 78% accuracy. The overall system is divided into three parts: - profit analysis, Crop recommender and Crop Sustainability predictor. This system will help the farmers to choose the right crop and will decrease the chances of crop failure and increase productivity.

Pradeepa B., et.al. [3] has proposed a paper in which they are making a recommendation system through integrated models of collecting environmental factors using Arduino microcontrollers and Machine learning techniques. In this model the data of environmental factors for plant growth is being collected and processed with the trained sub-models of the main of the system. The methodology of this system consists of several steps, first step would be the "Dataset Collection" in which the data of temperature, humidity, soil, pH, sunlight, and soil moisture for a plant growth are being collected. After collecting the Dataset, "Collecting Environment Factors" step will be done in which to collect the environment factors some of the controllers and sensors are being used which are Arduino microcontrollers, micro-controllers for temperature and humidity sensors, sunlight intensity sensor, soil moisture sensor, and soil pH sensor. In the third step the "crop prediction" method takes place, in which the best crop to cultivate with high probability of growing is being predicted by using machine learning algorithms, The algorithms are Naïve Bayes and Support vector machine. And in the next step, the monitoring and feedback part is done. The overall accuracy of the proposed system is more than 92%. As the time passes, the accuracy level would be increased up to 95%. This system will suggest the best crop for cultivation for maximum harvest based on the tests conducted and data collected.

Potnuru S. N., et.al, [4] has proposed a paper in which they are using advanced regression techniques like Kernel Ridge, Lasso and ENet algorithms to predict the yield and uses the concept of Stacking Regression for enhancing the algorithms to give a better prediction. This model focuses on minimizing error and obtaining better predictions. The performance metric used in this model is Root mean square root. When the models are applied individually, for ENet it was around 4%, Lasso had error around 2%, for Kernel Ridge it was around 1% and finally after stacking it was found around less than 1%. In this model, they are making a web application to predict the crop yield for farmers.

Ramesh M., et.al, [5] has proposed a paper in which they are implementing the crop selection method that will help the farmers select the best crop for cultivation and solve many agricultures and farmers' problems. A java application is created which will help to predict the crop yield rate. This application consists of three parts: - managing datasets, testing datasets and analyzing the datasets. Two machine learning algorithms are being used to build the model, one is Naïve Bayes, and another is K-Nearest Neighbour.

Renuka, et.al, [6] has proposed a system in which the focus is on predicting crop yield with the use of supervised learning algorithms. It is noted that this type of algorithm helps to build most accurate and effective model. It involves several steps for predicting the best crop yield for farmers. Firstly, the dataset is collected from several sources for at least 10 years. The data sets include soil dataset, rainfall dataset, and crop yield dataset. After that, the Preprocessing step is done in which inserting the missing values, the appropriate data range, and extracting the functionality is done. After that split the dataset into train and test set with a division ratio of 80% and 20%. In this model the machine learning algorithms are being used that are Decision tree with 99% accuracy, KNN with 58%, SVM_rbf with 58.9%, and SVM linear with 53% accuracy. Hence, in this the most accurate algorithm is the Decision tree. This system helps the farmers to reduce the problem faced in terms of crop yield and improve in terms of earning high profits and maximizing profits.

Shikha U., et.al, [7] has proposed a model in which they are providing a resolution in terms of technology via ESM algorithm, that will help to achieve the right way of predicting the crop yielding in a global as well as industry context. This model works on the Ensemble Framework which will improve its overall performance. The

performance of the model is evaluated by the coefficient of determination and root mean square error. The ensemble technique is found to be the best technique for crop yielding.

Nidhi H K., et.al, [8] has proposed a paper in which they are improving crop productivity by machine learning algorithms using assembling technique. This paper's main aim is to improve crop productivity and choose the right crop for cultivation. In this paper the assembling technique is used to build a model that combines the predictions of multiple models together and will give the best crop as a result for cultivation. This model is based on the soil specific type and characteristics with high accuracy. The machine learning algorithms that are used in this model are Random Forest, Naïve Bayes, and Linear SVM. The average classification accuracy obtained by combining the independent base learners is 99.91%.

Zeel D., et.al, [9] has proposed an intelligent model named as AgroConsultant, in which it intends to assets the farmers that which crop should be grow based upon the sowing season, his farm's geographical location, soil characteristics as well as environmental factors like temperature and rainfall. In this model, firstly they have collected the data from 'India Agriculture and Climate Data Set', to train the system. It covers the data of over thirty years. After that the Data is being preprocessed. Then they have used the Machine learning algorithms like Decision Tree, K-NN, Random Forest, and Neural Network to obtain the highest possible accuracy rate. All four algorithms are being compared individually and the best algorithm that have highest accuracy rate is being selected for the model. Decision Tree with 90.20% accuracy, K-NN with 89.78% accuracy, Random Forest with 90.43% accuracy and Neural Network with 91% accuracy were found.

Avinash K., et.al, [10] has proposed a model in which they have applied SVM classification algorithm, Decision Tree algorithm and Logistic Regression algorithm and they have founded that the SVM classification model gives the best accuracy rate that is 89.66% as compared to the other algorithm. So, they have took that algorithm and make the model that will help the farmers to suggest the best crop as well as to suggest the pest control techniques. They have taken the dataset that includes the data of Soil, pH, rainfall and Temperature on which the model has been trained and give the result under these conditions.

Detection Algorithms	Originated	Description	Applications	Merit	Demerits
Support Vector Machine (SVM)	SVM was founded in 1997 by Marshall Reavis.	Used for Classification and regression tasks.	Text and hypertext categorization, handwriting recognition, Bioinformatics.	Works well with even structured and unstructured data.	Long training time for large dataset, hard to visualize their impact.
Decision Tree	The First Decision Tree Regression was invented in 1963	Supervised Learning technique.	Used in business for analyzing customer data and making marketing decisions.	Simple to understand, solve decision related problem.	Overfitting issue, contains a lot of layers which makes it complex.
Naïve Bayes	Naïve Bayes was introduced in 1960s.	The main aim of Naïve Bayes is to process, analyze, and categorize data.	Credit Scoring, Medical Data classification, Text classification.	Suitable for solving multi-level prediction problems, work quickly and saves a lot of time.	Less suitable for real-world use cases, Faces the 'zero-frequency problem'.
Neural Network	Neural Network was introduced in 1954.	Imitate human brain activity to solve difficult problems.	Image recognition, speech recognition, machine translation.	Effective visual analysis, processing of unorganized data.	Hardware Requirement, Incomplete results, Data suitability.
K-Nearest Neighbour	The K-Nearest Neighbour algorithm was introduced in 1951.	Non-parametric, supervised learning classifier.	Pattern recognition, data mining, financial market predictions, intrusion detection.	Simple to implement, robust to the noisy training data, more effective if the data is large.	Always need to determine the value of K which may be complex some time, computation cost is high.

II. IMPLEMENTATION AND OUTCOMES

The outcomes of crop recommendation model and the steps that are taken during the implementation are shown below:

1) Collecting the Dataset and Analysing the Conditions

To make the model that we help to recommend the crop, firstly we need to have the datasets by which we can train our system for predict the result. So, here we have taken the data of temperature, pH, rainfall, humidity, N, K, and P. We have taken almost dataset pf Twenty-Two hundred crops having these value conditions. We have checked then it shown not contain any null value, or any null row and column. This dataset contains the almost 22 crops that will have been recommended by using these conditions. And then our next step is to train the data and predict the result from that.

2) Training and predicting the result

Regarding agriculture, as we know that it is the biggest sector all over the world. As we know that the agriculture sector consists of variety of vegetables, pulses, fruits and many more. So, it would be also most challenging sector to pick up the right crop for cultivation because the weather condition, and the soil conditions are being changed a lot. In this model we are going to predict the crop for the farmers which will give the high crop yielding and high cultivation rate. So, it would be depending on different factors like temperature, pH, rainfall, N, P, K, and humidity. That we have seen earlier in this paper, so the next step is to train our model so that it will give the most accurate result. For that we have taken the machine learning algorithms like Logistic Regression, Naïve Bayes, Support Vector Machine, Extra Trees, K-Nearest Neighbour, Decision Tree, Random Forest, Bagging, AdaBoost, and Gradient Boosting. We have calculated the accuracy rate of these machine learning algorithms and choose the best accurate rate algorithm that we give us the best recommended result. So, we have found out that the Random Forest with accuracy rate 99.31% is the best algorithms for our model. We trained out model using this algorithm. The accuracy chart and the accuracy rate are given below.

After Training the model, now we have to predict the result. For that as we have mentioned that we have taken the Random Forest algorithm, we will take the values from the user through the web UI and show the best recommended result to the user.

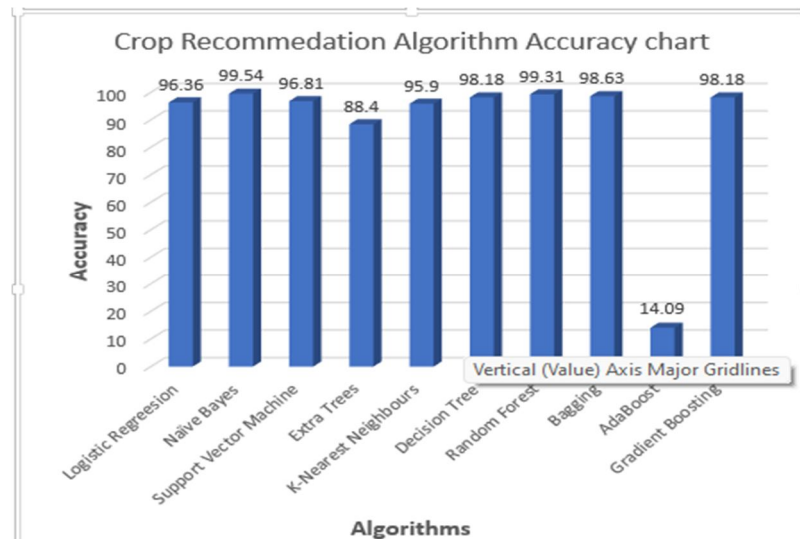


Figure. 1. Accuracy score of Algorithms used for recommendation system

The accuracy results of the different machine learning algorithms that would be used for crop recommendation system are shown below in Table 1.

TABLE I. ACCURACY SCORE OF DIFFERENT ALGORITHMS

Algorithms	Accuracy
Logistic Regression	96.36%
Naïve Bayes	99.54%
Support Vector Machine	96.81%
Extra Trees	88.40%
K-Nearest Neighbors	95.90%
Decision Tree	98.18%
Random Forest	99.31%
Bagging	98.63%
AdaBoost	14.09%
Gradient Boosting	98.18%

Figure. 2. User Interface for crop recommendation

III. CONCLUSIONS

In this study, we showed that the Machine Learning Algorithms is the Technique by which we can recommend the best crop for cultivation to the farmers. In this project, we have taken several machine learning algorithms and found out their accuracy rate, the best one we picked, and made our model to recommend the best crop under several conditions. We found out that the Random Forest would be the best algorithm with an accuracy rate of 99.31%. In the future, we will take the data in real-time so, that our system will become more accurate and also add a new feature like feedback by which we can take the feedback of farmers which helps us to make our model more accurate and recommend more accurate crops.

ACKNOWLEDGMENT

We owe a special debt of gratitude to our Professor Ms. Aditi Arora Department of Computer Science & Engineering for guiding us with her constant support and guidance throughout our work. We also take the opportunity to acknowledge the contribution of Professor (Dr.) Divya Mishra, Head, Department of Computer Science & Engineering for her full support and assistance during the development of the project.

REFERENCES

- [1] S. kanaga S. R., Rishi R., Sundaresan E., and Srijit V., "Demand Based Crop Recommender System for Farmers" in IEEE International Conference on Technological Innovations in ICT For Agriculture and Rural Development, 2020.
- [2] Priyadharshini A, Swapneel C., Ayush K., and Omen R. P., "Intelligent Crop Recommendation System using Machine Learning" in Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021).
- [3] Pradeepa B., Thilini W., Ruchirawya T. H., W.J.M. N., Dimantha M.A.C and Pabasara M.G.P., "Crop Recommendation System" in International Journal of Computer Applications (0975-8887), October 2020.
- [4] Potnuru S. N., Pinapa S. V., Bollu L. A., and B. Jabber, "Crop Yield Prediction based on Indian Agriculture using Machine Learning" in International Conference for Emerging Technology (INCET), Belgaum, India. Jun 5-7, 2020.
- [5] Ramesh M., Vijay S. R., and Shweta, "Crop Yield Prediction using Machine Learning Technologies" in 5th International Conference for Convergence in Technology (I2CT), Pune, India. Mar 29-31, 2019.
- [6] Renuka and Dr. Sujata T., "Evaluation of Machine Learning Algorithms for Crop Yield Prediction" in International Journal of Engineering and Advanced Technology (IJEAT), August 2019.
- [7] Shikha U., Pratima G., and S. Veenadhari, "A Crop Recommendation System to Improve Crop Productivity using Ensemble Technique" in International Journal of Innovative Technology and Exploring Engineering (IJITEE), February 2021.
- [8] Nidhi H K., Dr. G N S., Dr. B M S. and Dr. N K C., "Improving Crop Productivity Through a Crop Recommendation System Using Ensembling Technique" in 3rd IEEE International Conference on Computational Systems and Information Technology for Sustainable Solutions, 2018.
- [9] Zeel D., Subhash N., Rashi A., and Prof. Neepa S., "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms" in Fourth International Conference on computing Communication Control and Automation (ICCCUBEA), 2018.
- [10] Avinash K., Sobhangi S. and Chittaranjan P., "Recommendation System for Crop Identification and Pest Control Technique in Agriculture" in International Conference on Communication and Signal Processing, April 2019.