Crop Recommendation System using Machine Learning

Maaz Patel
Department of Computer Science Engineering
Devang Patel Institute of Advance Technology and Research
Charotar University of Science and Technology (CHARUSAT)
Anand, India
maazptl240602@gmail.com

Anagha Rane
Department of Computer Science Engineering
Devang Patel Institute of Advance Technology and Research
Charotar University of Science and Technology (CHARUSAT)
Anand, India
anagharane001@gmail.com

Vansh Patni
Department of Computer Science Engineering
Devang Patel Institute of Advance Technology and Research
Charotar University of Science and Technology (CHARUSAT)
Anand, India
20dcs093@charusat.edu.in

***Abstract*— Approximately 17% of India's GDP derives from the agricultural sector, which even employs more than 60% of the country's workforce. This field has seen some changes with new technologies like vertical farming and so on. However, many Indian farmers still follow traditional ways and beliefs to use their land. For example, they wait for the weather to match their farming practices, rather than adjusting to the weather changes. Our research goal is to help farmers pick the best crops for their situation and environment by predicting which crops fit well with the factors that influence crop growth, such as soil nutrients, soil pH, humidity and rainfall. We use different machine learning models, such as Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR), and Gaussian Naïve Bayes (GNB).**

***Keywords*— Crop suggestion, Nitrogen-Phosphorus-Potassium (NPK), Humidity, Rainfall, pH, Machine Learning (ML), Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR), and Gaussian Naïve Bayes (GNB).**

***Abbreviations*— Decision Tree (DT), Gaussian Naïve Bayes (GNB), Support Vector Machine (SVM), Logistic Regression (LR)**

I. INTRODUCTION

Agriculture is a significant area for the Indian economy and human survival. It is one of the primary occupations which is essential for human life. It likewise contributes a huge part to our day-to-day life [1]. In most cases, Farmers commit suicide due to production loss because they are not able to pay the bank loans taking for farming purposes[12]. We have noticed in present times that the climate is changing persistently which is harmful to the crops and leading farmers towards debts and suicide [18]. These risks can be minimized when various mathematical or statistical methods are applied to data and by using these methods, we can recommend the best crop to the farmer for his Agricultural land so that it helps him to get maximum profit [12].

In India today, agriculture has made significant advancements. Precision farming's secret weapon is "area-specific" cultivation. Although improvements have been made, there are still some problems with precision cultivation. Crop recommendations are significantly influenced by precision agriculture. Crop recommendations are determined by a variety of factors.

Precision agriculture focuses on identifying these parameters in an area-specific way to identify issues. Not all the results given by precision agriculture are accurate to result but in agriculture, it is significant to have accurate and precise recommendations because in case of errors it may lead to heavy material and capital loss. Many research works are being carried out, to attain an accurate and more efficient model for crop prediction [11].

Machine Learning focuses on the algorithm like supervised, unsupervised, and Reinforcement learning and each of them has its advantages and disadvantages. Supervised learning the algorithm assembles a mathematical model from a set of data that contains both the inputs and the desired outputs. An unsupervised learning-the algorithm constructs a mathematical model from a set of data that contains only inputs and no desired output labels. Semi-supervised learning-algorithms expand mathematical models from incomplete training data, where a portion of the sample input doesn't have labels [8].

This paper aims to recommend the most suitable crop based on input parameters like Nitrogen (N), Phosphorous (P), Potassium (K), PH value of soil, Humidity, Temperature, and Rainfall. This paper predicts the accuracy of the future production of twenty two different crops such as rice, maize, chickpea, kidney beans, pigeon peas, moth beans, mungbean, black gram, lentil, pomegranate, banana, mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, and coffee crops using various supervised machine learning approaches in of India and recommends the most suitable crop. The dataset contains various parameters like Nitrogen (N), Phosphorous (P), Potassium (K), PH value of soil, Humidity, Temperature, and Rainfall. This proposed system applied different kinds of Machine Learning algorithms like Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR), and Gaussian Naïve Bayes (GNB).

II. RELATED WORK

[1] Kumar, Y. Jeevan Nagendra, V. Spandana, V. S. Vaishnavi, K. Neha, and V. G. R. R. Devi. “Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector”. In this proposed system crop yield prediction can be done from the past historical data which includes factors such as temperature, humidity, ph, rainfall, crop name. Under this system, maximum types of crops will be covered across different districts of India. By applied this

proposed system, we can predicted best crop according to the field weather conditions. This crop prediction can be done by random forest algorithm and decision tree. By applying random forest algorithm got best accurate value result. More accuracy results gave more profit to the crop yield.

[9] Suresh, G., A. Senthil Kumar, S. Lekashri, and R. Manikandan. "Efficient Crop Yield Recommendation System Using Machine Learning For Digital Farming". This proposed system is used to identify particular crop according to given particular data. By applying Support Vector Machine (SVM) acquired higher precision and productivity. This research paper mainly worked on two datasets: sample dataset of location data and sample dataset of crop data. By using this proposed system recommended particular crop according to their Nutrients (N, P, K, and PH) values and also identified available Nutrients values and required fertilizers quantities for the particular crop like Rice, Maize, Black gram, Carrot and Radish.

[10] Reddy, D. Anantha, Bhagyashri Dadore, and Aarti Watekar. "Crop recommendation system to maximize crop yield in ramtek region using machine learning". This proposed system worked on three parameters: soil characteristics, soil types and crop yield data collection based on these parameters suggesting the farmer suitable crop to be cultivated. This proposed system worked on different machine learning algorithms like random forest, CHAID, K-Nearest Neighbour and Naïve Bayes. By applied this proposed system we can predict particular crop under particular weather condition, state and district values. Thus our proposed work would help farmers in sowing the right seed based on soil requirements to increase productivity of the nation.

[14] Rajak, Rohit Kumar, Ankit Pawar, Mitalee Pendke, Pooja Shinde, Suresh Rathod, and Avinash Devare. "Crop recommendation system to maximize crop yield using machine learning technique". This proposed method is used for identifying particular crop based on soil database. This proposed system worked on various crops like groundnut, pulses, cotton, vegetables, banana, paddy, sorghum, sugarcane, coriander and various attributes like Depth, Texture, Ph, Soil Color, Permeability, Drainage, Water holding and Erosion. This proposed system worked on various machine learning classifier like support vector machine (SVM) classifier, ANN classifier, Random Forest and Naïve Bayes for recommend a crop for site specific parameter with accuracy and efficiency. This research work would help farmers to increase productivity in agriculture, prevent soil degradation in cultivated land, and reduce chemical use in crop production and efficient use of water resources.

[15] Doshi, Zeel, Subhash Nadkarni, Rashi Agrawal, and Neepa Shah. "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms". In this research paper, developed an intelligent system called AgroConsultant. This proposed system can be divided into two sub-systems: i) crop suitable predictor ii) Rainfall Predictor. This proposed system are worked on five major (bajra, jowar, maize, rice and wheat) and fifteen minor (barley, cotton, groundnut, gram, jute, other pulses, potato, ragi, tur, rapeseed and mustard, sesame, soybean, sugarcane, sunflower, tobacco) crops and some attributes like Soil Type, Aquifer thickness, Soil PH, Thickness of topsoil, Precipitation, Temperature, Location parameters. In this proposed system Implemented different machine learning algorithms like Decision Tree, K Nearest Neighbor (K-NN), Random Forest and Neural Network and performed multi-label classification on it. This proposed system achieved 71% accuracy by using rainfall predictor model and achieved 91.00% accuracy by applying neural network algorithm on crop suitable predictor system.

[18] Dighe, Deepti, Harshada Joshi, Aishwarya Katkar, Sneha Patil, and Shrikant Kokate. "Survey of Crop Recommendation Systems". This proposed system developed a crop recommendation system for smart farming. In this research paper reviewed various machine learning algorithms like CHAID, KNN, K-means, Decision Tree, Neural Network, Naïve Bayes, C4.5, LAD, IBK and SVM algorithms. For this research used Hadoop framework for the intensive calculations and also helped to get better accuracy for the system.

[21] Kulkarni, Nidhi H., G. N. Srinivasan, B. M. Sagar, and N. K. Cauvery. "Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique". This proposed system is used for recommended the right crop based on the soil specific type and characteristics like average rainfall and the surface temperature with high accuracy. This proposed system worked on various machine learning algorithms like Random Forest, Naive Bayes, and Linear SVM. This crop recommendation system classified the input soil dataset into the recommendable crop type, Kharif and Rabi. By applying this proposed system achieved 99.91% accuracy result.

III. PROPOSED SYSTEM

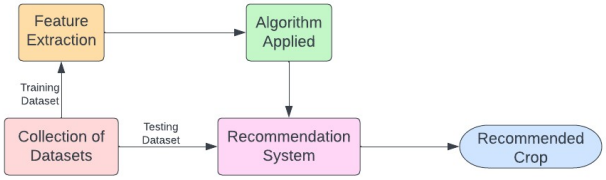

Fig 1: Block Diagram of Overall Methodology of Proposed System

In our framework, we have proposed a procedure that is separated into various stages as appeared in Figure 1.

The five phases are as per the following:
1) Collection of Datasets
2) Pre-processing (Noise Removal)
3) Feature Extraction
4) Applied Machine Learning Algorithm
5) Recommendation System
6) Recommended Crop

Flow of the Proposed System

As demonstrated in the figure, the methodology to extract the sentiment contains the several steps that are described below:

(1) *Data Collection:*

The dataset [27] consists of parameters like Nitrogen (N), Phosphorous (P), Potassium (K), PH value of soil, Humidity, Temperature and Rainfall. The datasets have been obtained from the Kaggle website. The data set has 2200 instance or data that have taken from the past historic data. This dataset include twenty two different crops such as rice, maize, chickpea, kidneybeans, pigeonpeas, mothbeans, mungbean, blackgram, lentil, pomegranate, banana, mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, and coffee.

The dataset is separated in Train and Test sets in which 80% of the whole dataset is taken as Train and 20% as Test dataset.

Dataset - Link:

https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset

(2) *Pre-Processing (Noise Removal):*

For the successful application pre-processing is required. The data which is acquired from different resources are sometime in raw form. It may contain some incomplete, redundant, inconsistent data. Therefore in this step such redundant data should be filtered. Data should be normalized [5]. We also use Power BI to remove peak/downfall, local min-max, outliers, and junk values.

(3) *Feature Extraction:*

This step is focus on identifying and using most relevant attribute from the dataset. Through this process irrelevant and redundant information is removed for the application of classifiers [5].

(4) *Methodology:*

In this proposed system applied different Machine Learning algorithms like Decision Tree, Support Vector Machine (SVM), Logistic Regression (LR), and GaussianNB.

A. Decision Tree:

Decision tree classifiers utilize greedy methodology. It is a supervised learning algorithm where attributes and class labels are represented using a tree [15]. The main purpose of using Decision Tree is to form a training prototype which we can use to foresee class or value of target variables by learning decision rules deduced from previous data (training data). The Decision tree can be described by two distinct types, namely decision nodes and leaves. The leaves are the results or the final end results. Each node in the tree acts as a test case for some attribute, and each edge descending from that node corresponds to one of the possible answers to the test case. This process is recursive in nature and is repeated for every sub-tree rooted at the new nodes [22].

We have applied Decision tree approach in our model as:

(i) Importing library DecisionTreeClassifier from sklearn.tree Class
(ii) Now we create DecisionTree Classifier object
(iii) In the last we fit our data

B. Support Vector Machine (SVM):

Support Vector Machine (SVM) is a supervised machine learning algorithm or model which can be utilized for classification and as well as for regression challenges. However, we mainly use it in classification challenges. SVM is generally represented as training data points in space which is divided into groups by intelligible gap which is as far as possible [22]. In SVM algorithm, each data item is plotted as a point in n-dimensional space with each feature value being the value of a specific coordinate. Then the classification is performed by finding the hyper-plane differentiating the two classes very well [19].

C. Logistic Regression (LR)

The Logistic Regression model is a broadly used statistical model that, in its basic form, uses a logistic function to model a binary dependent variable; many more complex extensions exist. In Regression Examination, Logistic regression is predicting the parameters of a logistic model; it is a form of Binomial regression [22].

D. Gaussian Naïve Bayes (GNB)

Gaussian Naive Bayes (GNB) is a classification technique used in Machine Learning (ML) based on the probabilistic approach and Gaussian distribution. Gaussian Naive Bayes assumes that each parameter (also called features or predictors) has an independent capacity of predicting the output variable. The combination of the prediction for all parameters is the final prediction that returns a probability of the dependent variable to be classified in each group. The final classification is assigned to the group with the higher probability.

IV. RESULT AND ANALYSIS

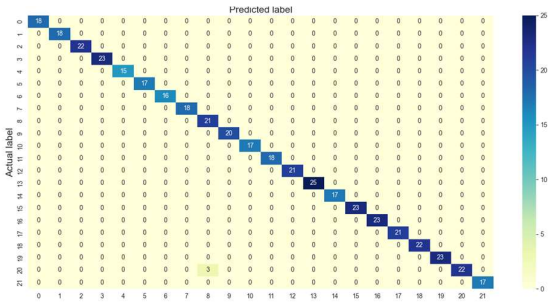

Fig 2 : Accuracy Comparison

The confusion matrix used to determine the performance of the classification models for a given set of test data. It can only be determined if the true values for test data are known.

The matrix itself can be easily understood, but the related terminologies may be confusing.

Training Accuracy Score	99.5%
Validation Accuracy Score	99.3%

Table 1 : Algorithm vice Accuracy Result in Percentage

Based on the results provided, we can see that the model performs great with 99.3% accuracy.

V. CONCLUSION

We have developed and applied a smart system that can suggest suitable crops for farmers across India. This system would help the farmers choose the best crop based on factors like Nitrogen, Phosphorous, Potassium, PH Value, Humidity, Temperature, and Rainfall. By using this research we can increase productivity of the country and produce profit out of such a technique. This research can enhance the country's productivity and profitability by using this technique. This way, farmers can grow the right crop and increase their income and the country's overall earnings. We have evaluated machine learning algorithms and discovered that Decision Tree and Gaussian NB had the best accuracy among them.

VI. FUTURE WORK

The system can be enhanced further to add following functionality:

1. The main future work's aim is to improved dataset with larger number of attributes.
2. We need to build a model, which can classify between healthy and diseased crop leaves and also if the crop has any disease, predict which disease is it.
3. To build website and mobile app for easy to use.

VII. REFERENCES

[1] Kumar, Y. Jeevan Nagendra, V. Spandana, V. S. Vaishnavi, K. Neha, and V. G. R. R. Devi. "Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector." In 2020 5th International Conference on Communication and Electronics Systems (ICCES), pp. 736-741. IEEE, 2020.

[2] Nigam, Aruvansh, Saksham Garg, Archit Agrawal, and Parul Agrawal. "Crop yield prediction using machine learning algorithms." In 2019 Fifth International Conference on Image Information Processing (ICIIP), pp. 125-130. IEEE, 2019.

[3] Medar, Ramesh, Vijay S. Rajpurohit, and Shweta Shweta. "Crop yield prediction using machine learning techniques." In 2019 IEEE 5th International Conference for Convergence in Technology (I2CT), pp. 1-5. IEEE, 2019.

[4] Jain, Sonal, and Dharavath Ramesh. "Machine Learning convergence for weather based crop selection." In 2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS), pp. 1-6. IEEE, 2020.

[5] Gandge, Yogesh. "A study on various data mining techniques for crop yield prediction." In 2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), pp. 420-423. IEEE, 2017.

[6] Suresh, A., P. Ganesh Kumar, and M. Ramalatha. "Prediction of major crop yields of Tamilnadu using K-means and Modified KNN." In 2018 3rd International Conference on Communication and Electronics Systems (ICCES), pp. 88-93. IEEE, 2018.

[7] Kamatchi, S. Bangaru, and R. Parvathi. "Improvement of Crop Production Using Recommender System by Weather Forecasts." Procedia Computer Science 165 (2019): 724-732.

[8] Bondre, Devdatta A., and Santosh Mahagaonkar. "Prediction of Crop Yield and Fertilizer Recommendation Using Machine Learning Algorithms." International Journal of Engineering Applied Sciences and Technology 4, no. 5 (2019): 371-376.

[9] Suresh, G., A. Senthil Kumar, S. Lekashri, and R. Manikandan. "Efficient Crop Yield Recommendation System Using Machine Learning For Digital Farming." International Journal of Modern Agriculture 10, no. 1 (2021): 906-914.

[10] Reddy, D. Anantha, Bhagyashri Dadore, and Aarti Watekar. "Crop recommendation system to maximize crop yield in ramtek region using machine learning." International Journal of Scientific Research in Science and Technology 6, no. 1 (2019): 485-489.

[11] Pudumalar, S., E. Ramanujam, R. Harine Rajashree, C. Kavya, T. Kiruthika, and J. Nisha. "Crop recommendation system for precision agriculture." In 2016 Eighth International Conference on Advanced Computing (ICoAC), pp. 32-36. IEEE, 2017.

[12] Garanayak, Mamata, Goutam Sahu, Sachi Nandan Mohanty, and Alok Kumar Jagadev. "Agricultural Recommendation System for Crops Using Different Machine Learning Regression Methods." International Journal of Agricultural and Environmental Information Systems (IJAEIS) 12, no. 1 (2021): 1-20.

[13] Jejurkar Siddhi, S., S. Bhosale Meghna, and D. N. Wavhal. "Crop Predication and Diseases Detection Using Machine Learning." (2021).

[14] Rajak, Rohit Kumar, Ankit Pawar, Mitalee Pendke, Pooja Shinde, Suresh Rathod, and Avinash Devare. "Crop recommendation system to maximize crop yield using machine learning technique." International Research Journal of Engineering and Technology 4, no. 12 (2017): 950-953.

[15] Doshi, Zeel, Subhash Nadkarni, Rashi Agrawal, and Neepa Shah. "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms." In 2018 Fourth International Conference on

Computing Communication Control and Automation (ICCUBEA), pp. 1-6. IEEE, 2018.

[16] Mahule, Ankit Arun, and A. J. Agrawal. "Hybrid Method for Improving Accuracy of Crop-Type Detection using Machine Learning." International Journal 9, no. 2 (2020).

[17] Akshatha, K. R., and K. S. Shreedhara. "Implementation of machine learning algorithms for crop recommendation using precision agriculture." International Journal of Research in Engineering, Science and Management (IJRESM) 1, no. 6 (2018): 58-60.

[18] Dighe, Deepti, Harshada Joshi, Aishwarya Katkar, Sneha Patil, and Shrikant Kokate. "Survey of Crop Recommendation Systems." (2018).

[19] Jaiswal, Sapna, Tejaswi Kharade, Nikita Kotambe, and Shilpa Shinde. "Collaborative Recommendation System For Agriculture Sector." In ITM Web of Conferences, vol. 32. EDP Sciences, 2020.

[20] Ghadge, Rushika, Juilee Kulkarni, Pooja More, Sachee Nene, and R. L. Priya. "Prediction of crop yield using machine learning." Int. Res. J. Eng. Technol.(IRJET) 5 (2018).

[21] Kulkarni, Nidhi H., G. N. Srinivasan, B. M. Sagar, and N. K. Cauvery. "Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique." In 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), pp. 114-119. IEEE, 2018.

[22] Kumar, Avinash, Sobhangi Sarkar, and Chittaranjan Pradhan. "Recommendation system for crop identification and pest control technique in agriculture." In 2019 International Conference on Communication and Signal Processing (ICCSP), pp. 0185-0189. IEEE, 2019.

[23] S. M. PANDE, P. K. RAMESH, A. ANMOL, B. R. AISHWARYA, K. ROHILLA and K. SHAURYA, "Crop Recommender System Using Machine Learning Approach," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 1066-1071, doi: 10.1109/ICCMC51019.2021.9418351.

[24] Bhuvan, S., Alankriti Jain, and J. Sanjeetha. "DigiFarm–A Machine Learning based holistic Crop Prediction Platform for Farmers." NVEO-NATURAL VOLATILES & ESSENTIAL OILS Journal| NVEO (2021): 11112-11129.

[25] Qiao, Mengjia, et al. "Crop yield prediction from multi-spectral, multi-temporal remotely sensed imagery using recurrent 3D convolutional neural networks." International Journal of Applied Earth Observation and Geoinformation 102 (2021): 102436.

[26] Meshram, Vishal, et al. "Machine learning in agriculture domain: A state-of-art survey." Artificial Intelligence in the Life Sciences 1 (2021): 100010.

[27] Dataset - Link:

https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset