Section A-Research paper
ISSN 2063-5346

Optimizing Crop Production: An Agronomic Advisor Application Based on Soil Nutrients

Pratik Harde, Ibin Babu, Prof. Prachi Patil, Prof. Monali Shetty

Computer Engineering, Fr. CRCE, Mumbai, 400050, Maharashtra, India.

edu.pratik21@gmail.com, ibin.p12@gmail.com, prachi@fragnel.edu.in, shettymonalin@gmail.com

ABSTRACT

Agriculture sector works as the source of raw material for non-Agricultural sectors. As much as 60% of the land is used for farming in India. It feeds around total of 1.2 billion in population. The population has also been increasing day to day and also the agriculture is not able to meet the demanded requirements for the increasing population. Crop yields, meanwhile, have already begun to suffer as a result of climate change. Unnatural climate changes can have a negative impact on food production and forecasting, which in turn can affect farmers' economics by resulting in low yields. Droughts, floods, heat waves, storms, and various more extreme weather events can destroy crops and reduce harvests, and climate change can bring about these conditions. Crop output may also be impacted by variations in temperature and rainfall patterns, which may change when crops are planted and harvested. These changes can make it difficult for farmers to predict their yields accurately, leading to economic losses. It is possible to produce precise crop prediction results by using the right parameters, such as soil nutrients properties (Nitrogen, Phosphorous, and Potassium, nutrients concentration, soil type and pH value), rainfall patterns, temperature patterns, soil structures, and other factors, such as crop diseases. A crucial aspect of agriculture is determining the best crop to grow, and in recent years, machine learning algorithms have become increasingly important in this process. This unique research is use of ML algorithms to better precisely recommend the crops based on the location. Supervised learning classification was used for the recomme ndations in this study. This study's main goal is to identify the most effective feature selection and classification techniques to predict the best harvest that will thrive in a particular environment, including temperature, rainfall, and geographic location in a given state, soil properties, including phosphorus (P), potassium (K), nitrogen (N), and pH value, as well as soil type. To provide recommendations for crops that are likely to thrive in a particular environment based on the available soil nutrients, an agronomic advisor application can be developed. The application can use a suitable classification algorithm to identify the most relevant features of the soil and environment. This can help farmers optimize their crop production and maximize their yields.

Keywords: Machine Learning, Agriculture, Soil nutrients, temperature patterns, Crop Recommendation, Random Forest, SVM, Decision Tree, Logistic Regression.

1 INTRODUCTION

Choosing the appropriate crop is a critical decision for farmers since it has a significant impact on the final yield and is Influenced by factors such as the environment and soil type. Selecting the right crop for a particular farm is a challenging choice that affects the yield. Expert advice on crop selection or recommendations can be time-consuming and expensive, making it difficult for many farms to afford it. Traditional methods of crop selection, such as expert consultations or field trials, can be costly and require significant investments of time and resources. As a result, many farmers may not have access to expert advice, which can limit their ability to optimize their crop production and maximize their yields. The use of machine learning algorithms and agronomic advisor applications can provide a more cost-effective and efficient alternative to traditional methods of crop selection. By analyzing relevant data and identifying the most important features, these applications can provide farmers with recommendations for the crops that are most likely to thrive in their specific environment, based on the available soil nutrients and other relevant factors.

The management of system crops to maximize agricultural productivity is one of the key areas of precision agriculture. Suggesting suitable crops based on data analysis can help increase crop production while minimizing resource usage, by identifying the most appropriate options from a dataset. These programs are crucial for decision-making because they assist users in maximizing gains or reducing risks.

It is vital to develop a system that might provide Indian farmers with predicted information so they could make informed crop decisions. In light of this, we propose a system, an intelligent system, that, before advising the user on the crop that would grow the best, would evaluate soil characteristics (N, P, K, soil type, pH value, and nutrients concentration), as well as environmental variables (rainfall, temperature, and geographic location in relation to state.

2 LITERATURE SURVEY

In paper [1], Professor Rakesh Shirsath and a number of coauthors suggested a system that helps users select the crop to be planted. The method is an online system that any registered farmer can access through a subscription and receive personalized information. The system has a module that compiles data from many sources on crops that have already been cultivated and recommends a crop that would be a match for planting. To make the procedure easier overall, artificial neural networks are used. In case the farmer has any issues while using the system, a feedback method is offered at the end to allow the developer to make any necessary improvements.

In their research [2], Ji-chun Zhao and Jianxin Guo consider knowledge databases to be big data and make conclusions from the data. The various modules considered include users, knowledge engineers, domain experts, humanmachine interfaces, inference engines, and knowledge banks. The knowledge acquisition system collects data for the decision system and creates a usable knowledge base to address the problem. The essay makes use of several Hadoop modules to extract features. It utilizes unstructured data, processes it with Hive, Mahout and NoSQL then stores the outcomes in HDFS. Only the data for the wheat crop was reported; other crops were not taken into account.

As mentioned in the paper [3], the RSF is a farming recommendation system that takes into account a data analysis, location detection module and crop growth database, storage module, and physiographic database. The related location discovery module identifies areas nearby the user's current location and researches the crops that are grown there. As a result, recommendations are offered to the user using a similarity matrix. The location detection module uses the Google API services to determine the user's current location and identify similar sites that are close by. Nevertheless, the system does not get user feedback to improve the process.

The system suggested in paper [4] by authors S. Pudumalar and associated co-authors uses an ensemble technique known as Majority Voting Technique, which taps the power of several models to improve prediction accuracy. The final forecast is accurate when the majority voting mechanism is used, even if one of the methods predicts incorrectly. KNN, Random Trees, CHAID, and Nave Bayes for ensemble are the techniques used. The key elements used in the prediction process are if-then rules. The ensemble model provided 88% accuracy.

The research by Yogesh Gandge and Sandhya [5] is a review paper that looks at several algorithms and how effective they are for use in agriculture. It was discovered that multiple linear regression offered a rice yield accuracy of 90–95%. The ID3 algorithm was used to study the decision tree and generate suggestions for the soybean crop. The third method, SVM, was applied to all the crops and had good accuracy while utilizing little computer power. A neural network was used to corn-related data to obtain 95% accuracy. Also used were LAD Tree, K-means, KNN, C4.5, J48 and Naive Bayes. The investigation came to the conclusion that the algorithms still needed to be enhanced for greater accuracy. A dataset from Kaggle.com was analyzed for a study titled Agricultural Yield Prediction using Data Mining [6]. The LAD Tree, J48, LWL, and IBK algorithms were utilized by the author to analyses the data using the WEKA tool. The accuracy was evaluated using specificity, accuracy, RMSE, mean absolute error and sensitivity. Confusion matrices were used to find the situations that each classifier correctly identified. The results suggested that pruning could lead to improved accuracy.

In their study [7] recommended employing ANN, KNN, SVM, GBDT, Random Forest, Decision Tree and Regularized Gradient Forest as seven machine-learning approaches for crop selection. The system is designed to recover every crop that was sown as well as the timing of its growth at a particular time of the year. The crops providing the best yields are selected once the yield rate of each crop has been determined. In order to have the best yields, the approach also recommends which crops should be planted in what order.

3 METHODOLOGY

The most appropriate crop will be suggested using the proposed approach, which will make use of many soilrelated characteristics. The suggested system's technique consists of multiple steps, as indicated in Fig.1.

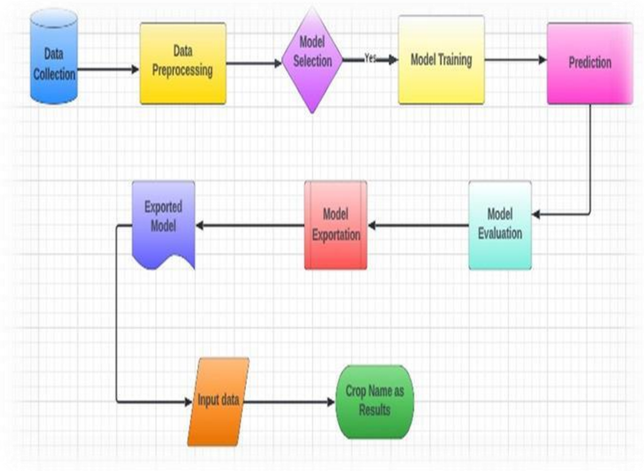

Fig.1 Proposed Architecture Diagram The process includes:

Data collection: This involves gathering data on soil-related parameters, such as soil pH, nutrient content, and other relevant information.

Data preprocessing: The collected data is then processed and cleaned to ensure that it is accurate and suitable for analysis.

Model selection: Based on the characteristics of the problem and the data at hand, the best machine learning model is chosen.

Model training: The selected model is then trained using the preprocessed data to create a predictive model.

Prediction: After being trained, the model might be used to forecast the optimal crop based on the input data.

Model evaluation: The performance of the model is then evaluated to make sure that it is accurate and reliable.

Input data: The input data includes soil-related parameters and other relevant information that is used to generate recommendations.

Crop name as an outcome: Based on the input data, the system's final output is the suggested crop name.

Overall, this process utilizes machine learning algorithms to analyze soil-related parameters and generate personalized recommendations for suitable crops. Farmers can use the agronomic advisor application to determine what crops to grow in a particular environment.

3.1 Dataset Collection

The process of building a machine learning model begins with collecting data. It is important to collect data before constructing a machine learning model. It is crucial to obtain a sizable amount of reliable information that is pertinent to the issue at hand. Data collection allows us to keep track of past events, which allows us to use data analysis to find repeated patterns. This dataset covers characteristics specific to the soil. This primarily consists of the soil and agricultural dataset for India over the preceding 20 years. This information collection consists of over 2000 observations from over 20 distinct crops. In essence, crops rest on the ground. The soil's fertility is influenced by its texture and the presence of nutrients like nitrogen, phosphorus, and potassium. This information also includes historical weather patterns for India.

1	N	P	K	temperatu	humidity	ph	rainfall	label
2	90	42	43	20.87974	82.00274	6.502985	202.9355	rice
3	85	58	41	21.77046	80.31964	7.038096	226.6555	rice
4	60	55	44	23.00446	82.32076	7.840207	263.9642	rice
5	74	35	40	26.4911	80.15836	6.980401	242.864	rice
6	78	42	42	20.13017	81.60487	7.628473	262.7173	rice
7	69	37	42	23.05805	83.37012	7.073454	251.055	rice
8	69	55	38	22.70884	82.63941	5.700806	271.3249	rice
9	94	53	40	20.27774	82.89409	5.718627	241.9742	rice
10	89	54	38	24.51588	83.53522	6.685346	230.4462	rice
11	68	58	38	23.22397	83.03323	6.336254	221.2092	rice
12	91	53	40	26.52724	81.41754	5.386168	264.6149	rice
13	90	46	42	23.97898	81.45062	7.502834	250.0832	rice
14	78	58	44	26.8008	80.88685	5.108682	284.4365	rice
15	93	56	36	24.01498	82.05687	6.984354	185.2773	rice
16	94	50	37	25.66585	80.66385	6.94802	209.587	rice
17	60	48	39	24.28209	80.30026	7.042299	231.0863	rice
18	85	38	41	21.58712	82.78837	6.249051	276.6552	rice
19	91	35	39	23.79392	80.41818	6.97086	206.2612	rice
20	77	38	36	21.86525	80.1923	5.953933	224.555	rice

3.2 Pre-processing

Pre-processing entails taking away anomalies and disturbances from the CSV dataset. Data loss frequently occurs and has a direct impact on the final machine learning model's effectiveness and accuracy. This needs to be addressed using a variety of strategies, including median and overall column mean. We can quickly clean the dataset using skLearn. It offers an imputer class that addresses and replaces values when they are absent. The imputer class accepts arguments such as missing values and techniques the imputer employs internally. Using the mean on-axis, missing data are renewed using this method.

3.3 Feature Selection

To provide recommendations for suitable crops based on soil related parameters, the following features can be considered:

Nitrogen (N): Nitrogen is a essential nutrient that is vital for plant increase, and it performs a critical role in the improvement of plant life. The availability of nitrogen in the soil can influence the growth and yield of crops. Therefore, the nitrogen level in the soil can be a crucial feature for crop recommendation.

Phosphorus (P): Phosphorus is another crucial nutrient for plant growth and it is involved in various plant processes, such as photosynthesis, energy transfer, and root development. The availability of phosphorus in the soil can significantly impact crop growth and yield.

Potassium (K): Potassium is a vital nutrient that is involved in several plant processes, including water regulation, photosynthesis, and disease resistance. Therefore, the availability of potassium in the soil can be an important feature for crop recommendation.

Temperature: Temperature is critical environmental aspect which could impact crop growth and improvement. Different crops have different temperature requirements for optimal growth and yield. Therefore, temperature may be a critical feature in figuring out the most appropriate crop for a particular region.

Rainfall: Rainfall is some other critical environmental element which could impact crop growth and improvement.
Different crops have different water requirements, and the availability of rainfall can significantly influence the growth and yield of crops. Therefore, rainfall can be an essential feature for crop recommendation.

Soil pH: Soil pH can have a significant impact on plant growth and development. Different crops have different pH requirements, and the availability of soil nutrients can be influenced by the soil pH. Therefore, soil pH can be a crucial feature for crop recommendation.

Humidity: Humidity is every other crucial environmental aspect which could affect crop growth and development. . Different crops have different humidity requirements for optimal growth and yield. Therefore, humidity can be a critical feature in determining the most suitable crop for a
specific region.

To provide farmers with informed recommendations about what crops to grow based on their region's soil type and environmental conditions, the crop recommendation system takes into account various soil-related parameters as features. This allows farmers to make informed selections about the quality crops to develop.

3.4 Choosing Machine Learning Model

When choosing a machine learning algorithm, Random Forest is one of the most liked and widely accepted supervised learning techniques. It constitutes a number of decision trees for differential subsets of data, other than using the whole data as a single unit. This helps to enhance the accuracy of the prediction of each variable. It makes sense to use Random Forest on a large dataset as it can provide results with high accuracy in a minimum span of time. Random Forest, Decision Tree, Logistic Regression and SVM are used in this model.

3.4.1 Decision Tree

Decision trees are a type of tool that predicts outcomes based on characteristics. They use simple rules to make predictions

Decision trees can handle both numerical and nominal data. The final prediction is made by gathering and combining all the results.

The Random Tree is a special type of decision tree that selects a random subset of characteristics to make predictions. The Random Tree builds multiple decision trees and combines their results for the best prediction

3.4.2 Random Forest

Random Forest helps to minimise overfitting in decision trees by introducing randomness in the tree construction process. The technique can handle missing data and still produce accurate results. The computational complexity of Random Forest is relatively low compared to other ensemble learning methods. Random Forest can be used for feature selection to identify the most important variables in a dataset.

The technique is insensitive to outliers and noisy data. Random Forest can be easily parallelized, Given that it is capable of handling datasets having continuous as well as categorical variables, the Random Forest is a flexible approach that may be utilized for a variety of applications. Random Forest often outperforms other machine learning algorithms when dealing with classification challenges.

The steps followed by the Random Forest algorithm are:

1. The original dataset is divided into n smaller bagged samples of size n.
2. A decision tree is constructed using input from all N bagged datasets. To calculate the ideal split, impurity measures such as Gini Impurity or Entropy are used, and M features are selected at random from the overall number of features in the training set to avoid looking at every feature in the dataset during a node split.
3. The unique outputs of each decision tree are combined into a single result.
4. For each observation, the outcomes produced by each tree are calculated, if working on a regression problem.
5. The majority vote is used to make a decision depending on the votes of the majority, if working on a classification problem.

3.4.3 Support Vector Machine (SVM)

SVM purpose is to identify a hyperplane in which the data points can be effectively separated (where N is the number of characteristics). In situations where there are two classes of data points, several hyperplanes can be utilized for classification. Data points are grouped into classes, and a hyperplane with the highest margin of difference is selected. By maximizing the margin distance, the accuracy of future data point classifications is enhanced.

ISSN 2063-5346

3.4.4 Logistic regression

Observations are categorised into distinct classes using a method known as logistic regression. Examples of categorization problems include whether something is spam or not, whether an online transaction is fraudulent or not, and if a tumour is malignant or benign. The output of logistic regression is converted into a probability value via the logistic sigmoid function.
Which logistic regression models are there?
1. Binary
2. Failure of the class of multilinear functions, such as sheep, dogs, or cats.
An approach for predictive analysis is logistic regression, a machine learning technique that is used for classification issues and is based on the probability notion.

3.5 Crop Prediction

To determine the best crop variety for a particular region, the crop recommendation system employs a machine learning algorithm that takes into account the unique environmental factors of the area. The system utilizes user input data to train the model and identify the crop with the highest probability of success. To determine the ideal crop type, machine learning techniques including SVM, RF, logistic regression, and decision trees are used. The technology evaluates variables like humidity, soil moisture, temperature, and pH levels to suggest which crops farmers should produce.

4 RESULT ANALYSIS

The proposed crop recommendation model relies on a crop database and soil factors to suggest the best crop for a given soil type. The best crop variety is identified using machine learning algorithms, and the system found that the Random Forest technique generated the most accurate results. Table 1 shows the accuracy rates for each algorithm evaluated by the system.

Table 1 Proposed methodology performance analysis

Algorithm	Accuracy
Logistic Regression	95%
Decision Tree	90%
Random Forest	99%
SVM	97%

In this field, a number of prototypes have been put out that are helping to solve agricultural problems. Indian agriculture has enormous untapped potential. The technology that will help farmers by giving them the necessary advice on crops, their growth, and other fundamental information still has to be improved to be more compact, accurate, and affordable. The majority of the approaches now in use involve manually determining the soil type. The approach had a number of drawbacks. The system might not offer the necessary support in other circumstances. Therefore, this work suggested a novel strategy that is based on the location, to recommend crops and other strategy to address the drawbacks of existing papers.

Table 2 Existing research papers performance analysis

Algorithm Used	Accuracy	Drawback	Citation
Subscription based system, ANN	94.5%	Earlier-planted Crops unknown to the system.	[1]
Inference engine ANN	90%	Can Hadoop and Artificial Neural Networks work together to produce better results.	[2]
Recommendation generation module.	92.4%	Crop growing season, crop output rate, and physiographic factors Database of seasonal crops.	[3]
Tools like CHAID, random tree, naive bayes KNN and WEKA	88%	Missing and out of range value.	[4]
Multiple Linear Regression, SVM, Decision tree utilising ID3, K-means, C4.5, Neural Networks and KNN	MLR= 90% NN=95%	For the algorithms to become more accurate, improvement is required.	[5]
J48, LAD tree, LWL, IBK algorithm	IBK gives the highest accuracy	The LAD tree displayed the least accuracy. The tree can be pruned to reduce errors.	[6]

5. CONCLUSION

The crop recommendation system has been effective in creating a model that can foretell which crops would do best given various environmental characteristics including topography, soil type, and climate. The system analyzes vast amounts of data to identify optimal crop varieties for different regions and soil types, considering factors such as temperature, rainfall, and soil fertility. After extensive research and analysis, the system has found that specific crops are better suited for particular regions and soil types. For example, crops that require high levels of rainfall are suitable for areas with high precipitation, while crops that can tolerate drought conditions are better suited for regions with low rainfall.

The predictive model developed by the system uses machine learning algorithms to analyze the data and provide customized recommendations to farmers. To recommend the best crops for a given region, the model consider a number of variables, including soil pH, nutrient content, and other environmental circumstances. By providing personalized recommendations, the system can help to increase crop yields, reduce costs, and improve efficiency in the agricultural sector. This, in turn, can contribute to sustainable agricultural practices and food security.

6. REFERENCES

1. Suruliandi, A., Mariammal, G., & Raja, S. P. (2021). Crop prediction based on soil and environmental characteristics using feature selection techniques. Mathematical and Computer Modelling of Dynamical Systems, 27(1), 117-140.

2. Doshi, Z., Nadkarni, S., Agrawal, R., & Shah, N. (2018, August). 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.

3. S. P. Raja, B. Sawicka, Z. Stamenkovic and G. Mariammal, "Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers," in IEEE Access, vol. 10, pp. 23625-23641, 2022

4. Kulkarni, N. H., Srinivasan, G. N., Sagar, B. M., & Cauvery, N. K. (2018, December). 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.

5. Pudumalar, S., Ramanujam, E., Rajashree, R. H., Kavya, C., Kiruthika, T., & Nisha, J. (2017, January). Crop recommendation system for precision agriculture. In 2016 Eighth International Conference on Advanced Computing (ICoAC) (pp. 32-36). IEEE.

6. Pande, S. M., Ramesh, P. K., Anmol, A., Aishwarya, B. R., Rohilla, K., & Shaurya, K. (2021, April). Crop Recommender System Using Machine Learning Approach. (Pp. 10661071),IEEE Xplore.

7. Liying Yang (2011), 'Classifiers selection for ensemble learning based on accuracy and diversity' Published by Elsevier Ltd. Selection and/or peer-review under responsibility of [CEIS].

8. Eswari, K. E., & Vinitha, L. (2018). Crop yield prediction in Tamil Nadu using Baysian network. International Journal of Intellectual Advancements and Research in Engineering Computations, 6(2), 1571-1576.