Crop selection and yield prediction Using ML Techniques

Dr. Subhash Bhagavan Kommina[1], Dr.A.V.N.Chandra Sekhar[2],Srinadh Unnava[3], G.Nageswara Rao[4], T.Vinay[5]

Professor[1,2], Associate Professor [3,4], Assistant Professor [5]
Department of Information Technology,
Sasi Institute of Technology & Engineering, Tadepalligudem, India

Abstract- "One of the most effective means of reducing severe poverty and promoting shared prosperity is agricultural development. Compared to other sectors, the agriculture sector's growth is 2-4 times more successful at boosting the lowest-income people's incomes. But today, growth that is based on agriculture is under trouble. Crop yields may be even more reduced by accelerating climate change and picking the wrong crop, particularly in areas with the greatest food insecurity, as well as most of the farmers expect their crop with high yield without knowing the land fertility this leads to increase in suicide rate of the farmers. The goal of this project is to create a system for advising on the optimal crop by taking attributes like humidity, Rainfall and soil parameters into consideration and also forecasts agricultural yield by taking land area into account. These functions will be carried out using the best machine learning approaches, which will be chosen after comparing the outcomes of all ML techniques effective rainfall has a great impact on the crop's growth in agriculture."

Keywords-Precision agriculture, recommendation system, crop yield.

I. INTRODUCTION

One of the major occupations in the country is cultivation. Small farms are typical of India. Less than 5 acres make up more than 75% of the nation's total land area. [1,2] Only around 45.6% of the land is irrigated, hence the majority of crops are rain-fed. India's land is cultivated on 60.46% of the country's territory. By doing so, it helps 1.2 billion people get what they need.[3] Almost 48% of farmers would not wish their children to continue farming and instead they wish them to study and settle in cities. Farmers were committing suicides at a higher rate in every year. The reason for this is low crop production and poor crop selection decisions. and farmers spend huge amount for obtaining high yield This lead to loss of money and going into debts [5]. Most of the times. Using conventional and non-scientific approaches, farmers frequently fail to choose the optimal crop for their land, which presents a severe issue for a nation.[8] Hence Crop protection, land assessment, and crop yield forecasting are more important to food production. Each farmer makes an effort to learn about crop yield and determine whether it satisfies their expectations. Many studies on crop yield have been conducted, the majority of which are empirical studies in the agriculture industry. Crop yield is difficult to describe, because there are numerous elements that influence yield prediction [13].

A subfield of intelligence (AI) that has been since the 1950s and is currently experiencing growth is machine learning. ML has been more effective thanks to modern computer technology's ability to solve mathematical equations, which has increased its appeal. The main goal of ML research is to apply machine learning to practical problems by decomposing them into manageable components that may be handled independently using one or more machine learning algorithms [16]. Additionally, by enhancing productivity and streamlining the overall procedure, these technologies can benefit [2].

The primary purpose of this survey is to recommend crop before seed sowing and predicting yield of that crop using machine learning Techniques that integrate by using environmental data which are most required. This research has found that there are also some models which consider all the features used for determining the crop selection and yield prediction such that To suggest an appropriate crop to the user, environmental factors like rainfall, temperature, and geographic location in terms of the state are taken into account along with soil characteristics like soil type, pH value, and nutrient concentration[7,11].Crop production mostly depends on climate factors, soil quality, landscapes, pest infestations, availability of water, genotype, and planning of harvest activities, among other factors[4].

The mentioned features come under the most important categories required for crop recommendation and yield prediction. They are soil Information, Humidity, Nutrients, Field Management, Solar Info and other parameters.

1)Soil Information: Soil parameters plays important role for crop growth. Soil contents have to be rich in minerals. There are various parameters for each crop are different.

https://doi.org/10.5281/zenodo.14190838

Some of the parameters are Soil type, PH soil, The pH range of most soils is between 3.5 and 10. Natural pH of soils normally ranges from 5 to 7 in areas with more rainfall, and from 6.5 to 9 in dry regions. pH levels can be used to categorise soils Mineral also.

2)*Humidity:* This parameter is the value of water concentration present in the soil which depends on the pressure and temperature of the climate. It is also based on the Rain fall pattern and precipitation.

3)Nutrients: These are contents present in the soil which are essential for the growth and the plant immunity. Nutrients supplied by the soil are known as soil minerals namely like Organic Carbon, saturation percentage, Nitrogen and Phosphorous and potassium available, and etc.

4)Field Management: This management, which is also known as "land management," may concentrate on the variations in soil types and soil characteristics to identify particular interventions that are also intended to improve the soil quality for the chosen land use. This management is based on the supplements given to field like irrigation, fertilizers.

5)Solar Information: It is about the temperature and the area climatic condition and if there are radiations present in it or not.

6)*Other parameters:* There are some other parameters like wind speed, for example if the wind speed is more in that area most of the crop does give production.

The crops vary from different regions to different country. Prediction is done according to their regional dataset provided for the machine learning model. There are various other parameters are also play role in determining the yield of the different crop. some of them are NDVI, EVI, Rainfall, Humidity, Temperature, soil Texture, Wind velocity, Solar radiation, Pressure, Sulphur, Zinc, Boron, Calcium, Crop Information, Leaf Area Index, Phosphorus, Manganese, Organic carbon etc.

II. LITERATURE SURVEY

Rohit Kumar Rajak, et al [1] In this paper Author proposed precision agriculture, this approach is distinguished by the use of a soil database gathered from farms, crops supplied by agricultural specialists, and soil testing lab datasets to achieve metrics like soil. The soil testing lab data will be used by the recommendation system to gather data, execute an ensemble model with majority voting, and use SVM and ANN as learners to effectively and accurately recommend a crop for a site-specific parameter. Their future work will focus on an enhanced data set with numerous features and will also include yield prediction.

Y. Jeevan Nagendra Kumar, et al [2] In this paper Author designed a system like experienced farmer to predict the crop yield based on the whether conditions and soil conditions, which helps farmers. He described how Increment in the accuracy gives the increment in the yield profit where results are obtained from the Kaggle datasets. They built correlation plots for temperature, humidity and rainfall. by using machine learning algorithms. They compared various algorithms and predicted Random Forest algorithm gives best accuracy.

PANDE et al [3] proposed method that uses a mobile application to connect farmers. GPS assists in locating the user. The user enters the location and soil type. The most profitable crop list can be picked using machine learning algorithms, and they can also forecast crop yields for user-selected crops. A few Machine Learning algorithms, including Support Vector Machine (Classifier), Artificial Neural Network (ANN), Random Forest (RF), Multivariate Linear Regression (MLR), and K-Nearest Neighbor (KNN), are applied to forecast agricultural productivity. The Random Forest among them demonstrated the best outcomes with 95% accuracy. The algorithm also makes recommendations on when to apply fertilisers to increase yield.

Elavarasan Dhivya, et al [4] proposed crop productivity model, the suggested study builds The data parameters feed the recurrent neural network's progressively stacked layers. Based on the input factors, the Q-learning network creates a crop yield prediction system. The output values of the recurrent neural network are translated into Q-values via a linear layer. The reinforcement learning agent combines a threshold and a set of parametric variables that help forecast crop yield. The agent then obtains an overall score for the actions taken, which is determined by minimising error and maximising forecast with an accuracy of 93.7%, the suggested model accurately forecasts crop yield while outperforming other models and maintaining the original data distribution of 35 years data.

Sonal Agarwal et al [5] proposed a hybrid approach for crop yield prediction. In this research he worked on machine learning and deep learning techniques to predict crop yield. An experiment is done on a crop dataset by the proposed model. They took the parameters like atmospheric conditions, soil type and other soil parameters. They constructed a model in which required AI algorithms are utilized to provide best suitable crop to

https://doi.org/10.5281/zenodo.14190838

be grown on particular land. The implementation of the research work gets started by loading the dataset of crops that has been gathered. It is proceeded with importing the necessary libraries and packages and continued with performing data pre-processing. They compared two ensemble models DT+ANN+Random Forest and LSTM+RNN+SVM hence second model gives more accuracy.

Kanika Bhatnagar [6] In this paper, author determined crop's yield is done using data that has been recorded and that shows boundaries such as rainfall, temperature, pH, precipitation, and yield name. It gives us a picture of a much-anticipated plant that will fill the field according to the natural season. The application of new rural tactics demonstrates a reasonable understanding of horticulture. Machines that focus on algorithmic law like Random Forest and KNN (K Nearest Neighbor) rarely have these doubts. Before changes can be made, the main plant must be in fruit. The harvests were produced by the Random Forest techniques.

Priyadarshini, A., et al [7] has suggested a system which assists farmers in selecting the best crop by offering information that typical farmers do not consider, reducing the likelihood of crop failure and increasing output. Additionally, it keeps them from suffering losses. Author compared Decision Tree (81%) with K Nearest Neighbor (85%), Nearest Neighbor with cross validation (88%), Linear Regression Model (88.26%), Naive Bayes (82%), Neutral Network (89.88%), and Support Vector Machine (78%). They advocated adding a web interface and a mobile app in the future so that millions of farmers across the nation may receive the crop cultivation advice.

Firduos Hina et al [8] author has employed a vast dataset which includes all Indian states whereas the previous researches worked on a single state. They used Machine Learning techniques to predict the crop yield. They wanted to help the farmers in crop cultivation using Data mining. They took climate factors and crop quantity under consideration. The main method of this project is to pre-process the data supplied they used UI for Interaction. 14 Attributes are taken under consideration. They showed that some machine learning models are utilized more frequently.

Swapneel et al [9] Omen Rajendra observed farmers are not selecting best crop for their land, so they proposed a system which considers both environmental and soil parameters for recommending the crop. The reason behind this is a particular soil will support a crop whereas the weather conditions won’t support that, such that the yield will suffer. This system gives three results profit analysis, crop recommendation and crop sustainability prediction. They used many algorithms for this system, among them they decided neutral networks are the best.

M.A.Manivasagam, et al [10] proposed a model Environmental factor, such as soil ph and local temperature, may have an impact on crop productivity. Different processes have different parameters. The details of the region's crop, risk factors, and more information about crops in agriculture or several environmental science sectors. Various fields of information are obtained on the crop, and many analysts throughout the world have created various procedures and techniques. They suggest the crop based on Ensembling techniques. The model merely recommends soil type, and in accordance with soil type, it can recommend suitable crops. The final application is a web browser where we input an image of the soil, and it then forecasts the type of soil and, based on that, the sort of crop that is good for that soil.

Meghana S et al [11] worked on crop and yield prediction with fertilizer estimation. The main aim of their work is to create a one stop solution to various problems in the domain of agriculture. This project helps farmers choose the ideal crop for a higher yield by getting rid of the manual and imprecise methods they previously utilised. This method makes use of data from the Soil Resources Development Institute's soil series. They used Random Forest and Support vector Machine for this project. The main aim of our System is to Automate the current manual soil testing procedure. In this system, they built a handheld device using a pH meter which will give the pH value of soil. This system takes Ph and NPK values as input and predicts the crop based on particular soil and land /area using an IoT device and which is compared with master data.

Dhanush et al [12] proposed a crop prediction algorithm that uses weather variables to recommend the optimum crop for a specific piece of land. Their system will increase the profit and also avoids soil pollution. In this project they worked on predicting the Rainfall and crop. They used two algorithms for predicting rain fall and crop. Support Vector Machine for rainfall and Decision Tree for Crop. The main goal of their project is to reduce the losses incurred by farmers and to develop a farmer friendly system with GUI.

Madhuri Shripathi Rao et al [13] worked on Crop prediction using machine learning. Given the importance of crop prediction, several ideas have been put up in the past to increase crop prediction precision. In this study, feed-forward back propagation artificial neural networks have been used to anticipate and estimate different crop yields in rural areas based on soil factors (PH, nitrogen,

https://doi.org/10.5281/zenodo.14190838

potassium, etc.) and atmospheric parameters (rainfall, humidity, etc.) The algorithms are used in this are K-Nearest Neighbor Classifier, Decision Tree Classifier Entropy Criterion, Decision Tree Classifier Gini Criterion, Random Forest Classifier Entropy Criterion, Random Forest Classifier Gini Criterion the main conclusion is to determine the crop that would grow the most effectively on a given plot of land, a comparison of the three different supervised machine learning models (KNN, Decision Tree, and Random Forest) is conducted.

Mansing Rathod,et al [14] Agro-climatic input variables such soil characteristics, rainfall, and temperature generally have an impact on crop productivity. Agriculture depends heavily on predicting which crops will be suited for cultivation, and in recent years, machine learning algorithms have become increasingly important in this process. In this situation, we have created a recommendation system that forecasts the kind of crop that can be grown on a specific plot of land as well as its overall output.

Kevin Tom Thomas et al [15] In this paper Author has made a lot of progress towards agriculture sector. In order to obtain information about more crops, they used a vast dataset in their study. So it is possible to predict a greater variety of crops that can be cultivated in various soil conditions. Author compared a variety of machine learning models. In order to make accurate forecasts, we can choose the model that performs the best from those that are available. They built our project in such a way that it is easily accessible to all the f armers and with the advancement in technology we can incorporate more features into it. The algorithms are K-Nearest neighbour, Decision Tree, K Nearest Neighbour with Cross Validation, Naive Bayes, Support Vector Machine.

Jeevan Nagendra Kumar Y, et al [16] We used a variety of machine learning approaches in our proposed crop recommendation and yield prediction system. We utilised classification and regression algorithms to estimate the crop type and yield produced in quintals and hectares, respectively. Name of the state, name of the district, time of year, location, amount of rainfall, average humidity, and mean temperature would be the inputs. A distinct dataset with historical data—rather than user input—is provided for training and testing after data preparation and visualisation techniques.

Nischitha K et al [17] The author in this paper have created a farmer-friendly system with a graphical user interface (GUI) that will predict which crop would be the best fit for a specific piece of land and will also provide information about the necessary nutrients to add up, the necessary seeds for cultivation, the expected yield, and the market price. The algorithms are Rainfall prediction: -SVM algorithm, Crop prediction Decision tree algorithm by providing GPS coordinates for a piece of land and gaining access to the government's rain forecasting system, we can anticipate crops without gathering any additional data. We can also create a model to prevent food shortages and surpluses.

Janmejay Pant et al [18] Author proposed crop system, in order to forecast four popular yields, which are primarily grown throughout India, machine learning is applied. Once the crop yield for a certain location has been estimated, inputs like fertilisers may be administered differently based on the anticipated crop and soil needs. In this study, we create a trained model using machine learning techniques to find trends in the data, which is then applied to crop prediction. In this work, machine learning is used to forecast the yields of the four most widely grown crops in India. These crops include wheat, potatoes, paddy rice, and corn.

Zeel Doshi et al [19] Author presents an intelligent system, referred to as AgroConsultant wants to help Indian farmers choose their crops wisely by providing them with information, depends on the sowing season, the location of his farm, the type of soil, and environmental elements like temperature and rainfall. Five major crops (bajra, jowar, maize, rice, and wheat) and fifteen minor ones (barley, cotton, groundnut, gramme, jute, other pulses, potato, ragi, tur, rapeseed, and mustard, sesame, soybean, sugarcane, sunflower, and tobacco) are covered by the aforementioned parameters. Topsoil depth, pH of the soil, aquifer depth, precipitation, temperature, and location. In this they compared four models Decision Tree (90.20), K-NN (89.78), Random Forest (90.43), Neural Network (91.00) respectively. This has found that there are also other models which consider all the features used for determining the crop selection and yield prediction

II.ARCHITECTURE OF YIELD PREDICTION

In this survey, most of the papers used ML algorithms for predicting. This segment represent the systematically design of the model and it explains aspects in the process. This architectural diagram is conceptual view only.

The architecture for crop prediction involves the use of a machine learning algorithm that takes in various input parameters such as soil type, weather conditions, and past crop data, and predicts the appropriate crop to plant. The algorithm uses historical data on crop yields and soil

https://doi.org/10.5281/zenodo.14190838

conditions to make predictions on which crops are most likely to succeed in a given area. The model can be trained using a variety of machine learning algorithms such as decision trees, random forests, and gradient boosting machines.

For yield prediction, the architecture involves the use of a machine learning algorithm that takes in various input parameters such as weather conditions, soil type, and crop data, and predicts the expected yield for a given crop in a given area. The algorithm uses historical data on crop yields and soil conditions to make predictions on expected yields for a particular crop. The model can be trained using a variety of machine learning algorithms such as neural networks, support vector machines, and regression models.

In both cases, the architecture for crop prediction and yield prediction involves a number of steps, including data collection, data preprocessing, feature selection, algorithm selection, model training, and model evaluation. The accuracy of the predictions depends on the quality and quantity of the data used, as well as the

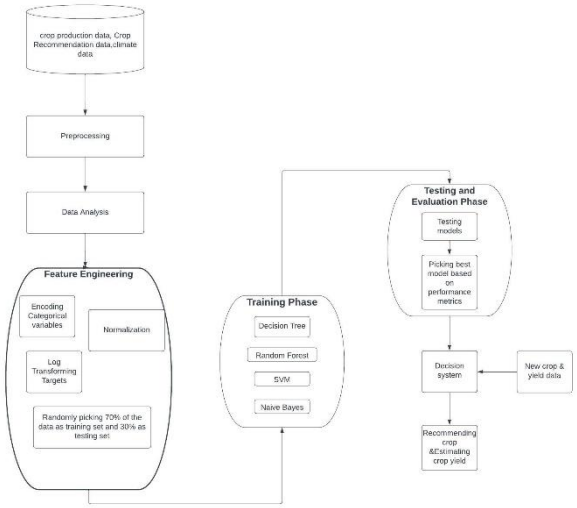
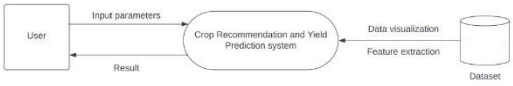


Fig 1: Architecture Diagram

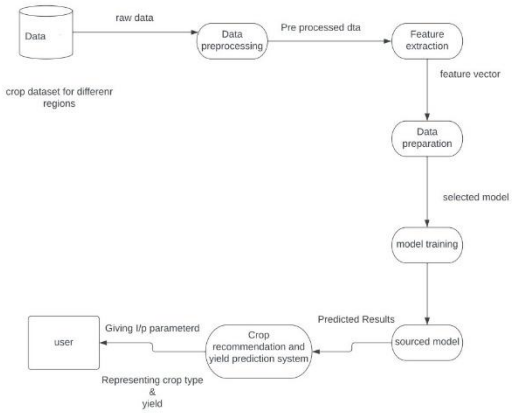
choice of machine learning algorithm and the model's parameters.Overall, the architecture for crop prediction and yield prediction involves the use of machine learning algorithms that leverage historical data and input parameters to make predictions on crop selection and expected yields. These models can be a valuable tool for farmers and other stakeholders in the agriculture industry, helping to improve crop selection, management, and planning, ultimately leading to increased yields and profitability.

The term "DFD" stands for "Data Flow Diagram," and it refers to a visual representation of the operations or procedures that collect, process, store, and distribute data across system components and between a system and its surroundings. There are no loops, decision rules, or control flows in a data flow diagram. At least one of the endpoints for each data flow. The visual representation makes it a useful tool for user and system designer communication. This DFD has three levels, which are categorised as levels 0, 1, and 2.

Fig: Level-0 DFD of System



This level is simple and user has to give the input



parameters and the predicting system flow of the system is described with detailed information. In this the user will give input parameters model. This saved model is ml algorithm which is will give output. These input

Fig: Level-1 DFD

parameters are based on the system required parameters only.

This DFD is an extended version of the above DFD, where the compared to other algorithm and gives better accuracy. The model is predicted by taking raw data into consideration. This raw data is in csv file format containing different datasets which undergoes pre-processing after data analysis. pre-processing is done for data cleaning which removes inconsistent data, adjusting null values, duplicate values, noisy data by different approaches. Feature extraction and selection is done by selecting the major impacted attributes taken from the data set. This can be done by correlation between features or by statistical tests or by recursive feature elimination.

https://doi.org/10.5281/zenodo.14190838

then the system predicts the results from the saved or by variance threshold. Then next step is algorithm selection. Data preparation is dividing the data into training and testing. This testing data validates the model and predicts the results. There are different techniques and

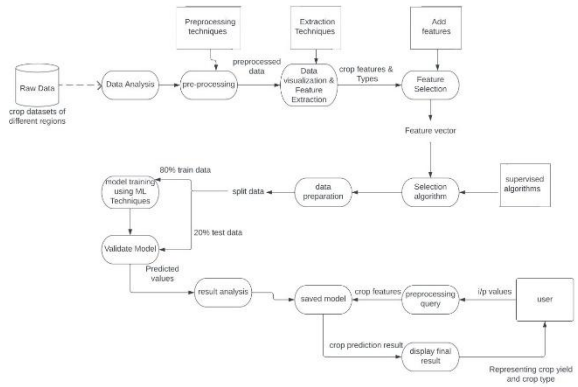


Fig: Level 2 DFD

approaches followed to perform This approach can give the results and predicts the results of crop recommendation of various crops for different regions and also can predicts the crop yield.

III.METHODOLOGY

A. *Techniques involved in existing systems*:

Random Forest: The supervised learning method includes the well-known machine learning algorithm Random Forest. It can be applied to ML Classification and Regression issues. The idea of ensemble learning serves as its foundation. In order to increase the dataset's predictive accuracy, the Random Forest classifier uses a variety of decision trees on different subsets of the input data.

Naïve Bayes: The Bayes optimistic Naive Bayes method is used. The possibility of likelihood is used to determine whether a specific component belongs to a specific class. Incredibly adaptable, naive Bayes classifiers demand different parameters directly in the number of elements in a learning issue.

SVM: One of the most well-liked algorithms for supervised learning is called the Support Vector Machine (SVM), and it is used to solve both classification and regression issues. It is largely utilised in Machine Learning Classification issues, though.

The K-NN algorithm makes the assumption that the new case and the existing cases are comparable, and it places the new instance in the category that is most like the existing categories. A new data point is classified using the K-NN algorithm based on similarity after all the existing data has been stored. This means that utilising the K-NN method, fresh data can be quickly and accurately sorted into a suitable category.

Decision Tree: It is a method that can be applied to classification and regression issues; however, it is most frequently used to address classification issues. This classifier has a tree structure.

Gradient Boost: It is a kind of ensemble learning method that combines a number of ineffective predictive models into a powerful one. As a result, a robust, highly accurate model that is resistant to overfitting is produced. A number of decision trees are fitted to the data as part of the algorithm's operation. A subset of the training data is used to construct each tree, and the output from each tree is merged with the output from the trees that came before it to get the final prediction. The approach estimates the gradient of the loss function with regard to the results of the prior trees at each iteration.

B. *Performance Analysis*

C. *Performance measures Accuracy*: Accuracy: This is the result of dividing the true positive and true negative values of a confusion matrix by the true positive, false positive, false negative, and true negative values.

$$Accuracy = TP + TN/TP + FP + FN + TN$$

Where TP stands for True Positive, TN for True Negative, FP for False Positive, and FN for False Negative. Precision: It is the result of dividing the True Positive value by the confusion matrix's total True Positive and False Positive values.

$$Precision = TP/TP + FP$$

https://doi.org/10.5281/zenodo.14190838

Precision means Where TP stands for True Positive and FP stands for False Positive.

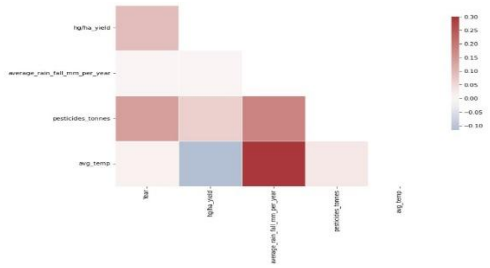


fig: Correlation of parameters for crop recommendation

III.EXPERIMENTAL SETUP

Crop Dataset: In this phase, data is collected on various crops and their yields from different regions. The data should include factors such as weather patterns, soil type, nutrient content, and other environmental variables that can affect crop yield. The data can be collected from different sources such as government agencies, research papers, Kaggle and online databases. A well-curated dataset is essential for the accuracy and effectiveness of the predictive models.

Uploading the dataset:

CSV data format is the most common format for ML data, but we need to take care about following major considerations while loading the same into our ML projects. In CSV data files, the header contains the information for each field. We must use the same delimiter for the header file and for data file because it is the header file that specifies how should data fields be interpreted. Uploaded Crop Recommendation.csv and Crop Production.csv which contains different parameters. There is a dataset used for this project, which is downloaded from Kaggle. The dataset Crop Recommendation.csv consists attributes namely N, P, K, Humidity, Rainfall, Temperature, Ph and other dataset crop production.csv consists of State Name, District Name, Crop Year, Season, Crop, Area, Production.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

+ Code + Text

```
[ ] df = pd.read_csv("/content/crop_production.csv")
    df[:5]
```

	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254.0	2000.0
1	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2.0	1.0
2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102.0	321.0
3	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176.0	641.0
4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720.0	165.0

Fig: sData Set

Looking at raw data:

It involves exploring the data in different ways, such as visualizations, statistical summaries, and descriptive analyses, to get a sense of what the data contains and what it can tell us. This process is an essential step in data analysis, as it helps to inform the selection of appropriate analysis methods and models, and to identify potential issues or challenges with the data. By looking at the dataset, we can also generate hypotheses and ideas for further investigation and analysis. The data set of crop recommendation contains crops like Apple, Banana, Black gram, Chick pea, coconut, Coffee, cotton, Grapes, Jute, Kidney beans,

Data Pre-processing: Once the dataset is collected, it needs to be pre-processed. This involves cleaning the data, removing any missing or duplicate values, and transforming the data into a format that can be used for analysis. The data may also need to be standardized or normalized to ensure that all variables are on the same scale.

Feature Selection: In this phase, the most important variables that affect crop yield are identified. This can be done using statistical methods or machine learning algorithms. The goal is to select a set of features that are most relevant to the problem at hand and can effectively predict crop yield.

Training Set and Testing Set: The dataset is split into a training set and a testing set. The training set is used to train the machine learning algorithms, while the testing set is used to evaluate the performance of the algorithms. The splitting ratio depends on the size of the dataset and the specific problem being addressed.

Model Selection: In this phase, the appropriate machine learning algorithm is selected to build the predictive model. Random forest, Naive Bayes, SVM, and decision tree are common algorithms used in crop yield prediction. The selection of the algorithm depends on the type of data, the problem at hand, and the performance metrics.

Model Training: The selected machine learning algorithms are trained on the training set to learn the relationship between the input variables and crop yield. This phase involves adjusting the parameters of the algorithm to optimize the model's performance.

Model Evaluation: The performance of the trained models is evaluated using various metrics such as accuracy, precision, recall, and F1 score. The testing set is used to evaluate the model's ability to generalize to new, unseen data. The model that performs the best is selected for further use.

Crop Selection: Using the trained model, the most suitable crop for a given region can be selected based on the predicted yield. This can help farmers optimize their crop selection and maximize their profits.

Yield Prediction: Finally, the trained model can be used to predict the yield of the selected crop based on the environmental variables in the region. This can help farmers plan their harvest and make informed decisions about their crops.

https://doi.org/10.5281/zenodo.14190838

IV.RELATED WORK

Title	Year	Algorithm	Results
[1] Crop recommendation system to maximize crop yield using machine learning technique.	2020	Naïve Bayes, random Forest, Support Vector machine, ANN.	-
[2] Supervised machine learning approach for crop yield prediction in agriculture sector.	2020	Random Forest, Decision Tree.	-
[3] Crop recommender system using machine learning approach	2021	ANN, Support Vector machine, Random Forest, KNN, Multi variate Linear Regression.	ANN 86%, Support Vector machine 75%, Random forest 95%, KNN 90%, Multi variate Linear Regression 60%.
[4] Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications.	2020	Proposed DRL algorithm, Deep learning algorithm, Artificial neural network Algorithm, Gradient boosting algorithm and Random forest algorithm, Bernoulli DBN, Bayesian ANN, Interval deep generative ANN.	Among compared algorithms Deep learning Reinforcement Algorithm gives best Accuracy of 93.7%.
[5] Hybrid approach for crop yield prediction using machine learning and deep learning algorithms	2021	SVM (Support Vector Machine)+LSTM(Long short term memory)+RNN (Recurrent Neural Network)compared with DT(Decision Tree)+ANN(Artificial Neural Network)+RNN.	SVM+LSTM+RNN gives 97% , DT+ANN+RNN gives 93% Accuracy.
[6] Agriculture Crop Recommendation System using Machine-Learning	2022	Random Forest,KNN.	Accuracy for Random Forest: 0.995 Precision for Random Forest: 0.996 Recall for Random Forest: 0.995 F1-score for Random Forest: 0.995 Accuracy for KNN: 0.988 Precision for KNN: 0.996 Recall for KNN: 0.989 F1-score for KNN: 0.989
[7] Intelligent Crop Recommendation System using Machine Learning.	2021	Decision Tree, KNN (K-Nearest Neighbour), Linear Regression, Naïve Bayes, Neural Network, Support Vector Machine.	Decision Tree gives 81%, KNN (K-Nearest Neighbour) gives 85%,Linear Regression gives 88.2%,Naïve Bayes82%,Neural Network gives 89.88%,Support Vector Machine 78%.
[8] Agriculture Crop yield Prediction Using machine learning	2022	Decision Tree	DT gives 95% Accuracy.
[10] An efficient crop yield prediction using Machine learning	2022	KNN,Decision Tree, Random forest.	KNN – 85%, Decision Tree – 90%, Random forest – 95%.

 https://doi.org/10.5281/zenodo.14190838

[11] Crop and Yield Prediction with Fertilizer estimation	2021	Compared Support Vector Machine, KNN, Random Forest Algorithms for both crop prediction datasets and for fertilizer also.	KNN gives accuracy 98.7% for crop prediction dataset, whereas for fertilizer Recommendation data set KNN gives 70% accuracy and Random Forest gives 98%.
[12] Crop prediction using machine learning approaches	2021	Decision Tree for crop prediction, SVM for Rainfall prediction.	-
[13] Crop prediction using machine learning	2022	KNN, Decision Tree.	K-Nearest Neighbor Classifier -97.04 Decision Tree Classifier Entropy Criterion 97.95, Decision Tree Classifier Gini Criterion - 98.86 , RandomForest Classifier Entropy Criterion 99.32 Random Forest Classifier Gini Criterion 99.32
[14] Crop Prediction Using Machine Learning	2018	Random Forest	Above 75%
[15] Crop Prediction Using Machine Learning	2020	KNN, Decision Tree, KNN with cross validation, Naïve Bayes, SVM.	KNN – 85%, Decision Tree – 81%, KNN with cross validation – 88%, Naïve Bayes – 82%, SVM – 78%.
[18] Analysis of agricultural crop yield prediction using statistical techniques of machine learning	2021	Decision Tree	DT gives the accuracy 96%.
AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms	2018	Decision Tree,KNN ,Random Forest,Neural Network.	Decision Tree gives 90.2%, KNN gives 89.78%, Random Forest gives 90.43%, Neural Network gives best accuracy 91%.

Table-1 Comparison Table for Algorithms

 https://doi.org/10.5281/zenodo.14190838

paper	nitrogen	p	k	calcium	PH	Humidity/moisture	rainfall	Temperature	location	Soil type	season	Land area	Crop name	other
1					Y					Y			Y	Permeability ,drinage, water holding, erosion,depth ,texture.
2					Y	Y	Y	Y					Y	
3							Y	Y	Y	Y	Y		Y	3
4	Y	Y	Y		Y	Y	Y	Y						Precipitation,density, electrical conductivity,wind speed,na,m
5					Y	Y	Y	Y				Y	Y	
6					Y	Y								Soil condition, weather,yield.
7	Y	Y	Y		Y		Y	Y	Y	Y			Y	
8							Y	Y		Y	Y		Y	Ground water
10	Y	Y	Y				Y			Y			Y	

Table-2 Different Parameters Used in Reference Papers

https://doi.org/10.5281/zenodo.14190838

paper	11	12	13	14	15	16	17	18	19
nitrogen	Y		Y		Y			Y	
p	Y		Y		Y			Y	
k			Y		Y			Y	
calcium	Y								
PH	Y	Y	Y		Y		Y		Y
Humidity/moisture	Y	Y			Y		Y		
rainfall		y		Y	Y	Y	Y	Y	Y
Temperature	Y	Y				Y	Y	Y	Y
location				Y					Y
Soil type									
season		Y							Y
Land area				Y					
Crop name	Y	Y	Y	y					
other	Electrical conductivity			Yield,msp, soil defeciencies .		Soil texture, yield		Fertiliser, rainfall, water flow.	Thickness of aquifier.

Table-3 Different Parameters Used in Reference Papers

Table-1 gives the comparison of referred papers of years (2018,2020,2021,2022). Most of the algorithms used in the papers are Decision Tree, Support vector machine algorithm, KNN, Random Forest for predicting the yield and crop selection.

Table-2, Table-3 contains comparison of the parameters that are taken to predict the crop yield. Most of the common parameters that are considered are nutrients (N, P, K, Ca), Rain fall , Temperature, humidity, PH , soil type etc, and other parameters considered in some referred papers are electrical conductivity , water flow, fertilizer used etc.

V. Result

In Prediction of crop and crop yield we are using SVM, Naïve Bayes,Decision Tree Classifier, Random Forest Classifier and logistic regression are used for the classification in our proposed system. We have collected the data and extract features for the output. After training the dataset and applying SVM, logistic regression, Naïve Bayes and Random Forest. Compare to other algorithms random forest algorithm gave accuracy of 99.09% which is very better accuracy from the previous works. Based on the input parameters of nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, and rainfall, a crop recommendation system was developed using the Random Forest algorithm. The model achieved an impressive accuracy score of 99.09%.This means that the model was able to accurately predict the appropriate crop to recommend based on the input parameters with a very high level of accuracy. This result suggests that the model can be relied upon to provide accurate and useful recommendations to farmers and other stakeholders in the agriculture industry.

Table: Accuracy Analysis for crop selection

Algorithm	Precision	Recall	Accuracy
Decision Tree	84.29	88.16	90
Naive Bayes	99	99	99.07
SVM	98.11	97.95	97.9
Logistic Regression	95.40	95.86	95.22
Random forest	99.29	98.9	99.09

The Random Forest algorithm was selected for this task because of its ability to handle a wide range of input data types and its high accuracy in prediction tasks. It is an ensemble learning algorithm that combines multiple decision trees to make a final prediction. By using a large number of trees, the model is able to capture complex patterns in the data and make accurate predictions even in the presence of noise and outliers.

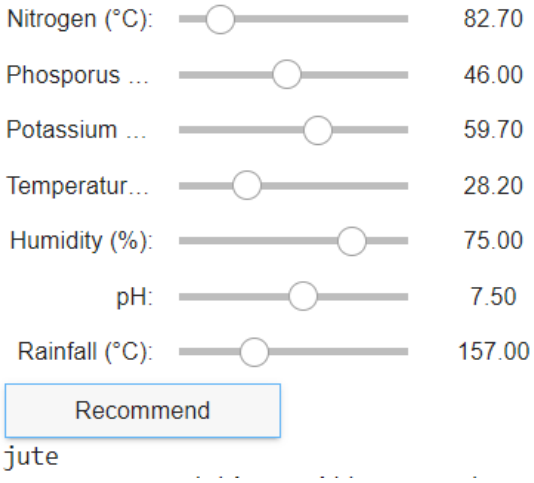


Fig : Widget User Interface

Overall, the high accuracy achieved by the Random Forest model in this crop recommendation task suggests that it has the potential to be a valuable tool for improving agricultural practices and increasing crop yields. The above figure describes the widgets user interface for crop recommendation.

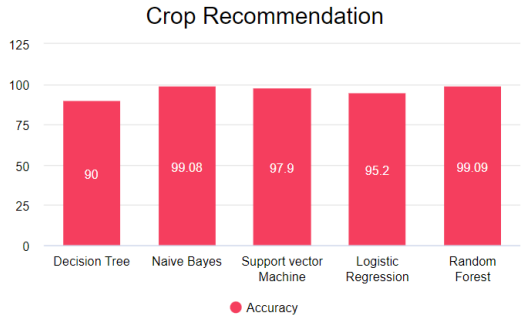


Fig:Results for crop recommendation

A crop yield prediction model was developed using a decision tree algorithm, using input parameters of State Name, District Name, Crop Year, Season, Crop, Area, and Production. The model achieved an accuracy score of 95.93%.

https://doi.org/10.5281/zenodo.14190838

This means that the model was able to accurately predict crop yields based on the input parameters with a high level of accuracy. The model's ability to accurately predict crop yields can be a valuable tool for farmers and other stakeholders in the agriculture industry, allowing them to make more informed decisions about crop selection, management, and planning.

Table: Accuracy Analysis for crop yield prediction

Algorithm	Precision	Recall	Accuracy
Decision Tree	90.50	94.51	95.93
Gradient boost Regressor	87.4	87.9	89.65
Random forest	67.8	68.09	68.42

The decision tree algorithm was selected for this task because of its ability to handle a mix of categorical and numerical data types and its interpretability, which allows users to easily understand how the model arrived at its predictions. The algorithm builds a tree-like model of decisions and their possible consequences, allowing for easy visualization and interpretation of the decision-making process.

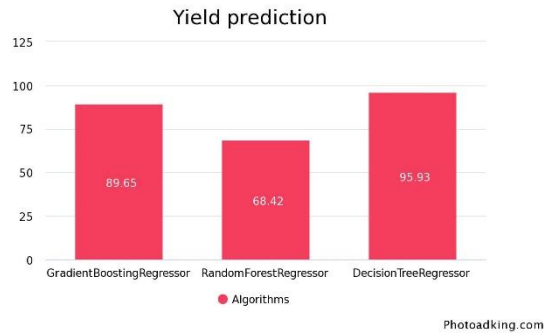


Fig: Results for yield prediction

The above graph depicts the comparison of algorithms for crop yield prediction which are gradient boost, Random Forest Regressor and Decision Tree Regressor with their accuracies.

VI.CONCLUSION

This paper provides a survey on the prediction of crop selection and yield prediction using different techniques. Different methods to predict is mentioned. Multiple techniques can be used such as SVM, Naive Bayer's, neural networks, etc. It is also seen in a few papers that models using neural networks give slightly better performance. This paper would like to mention that considered models did not include the parameter like wind velocity, soil texture, soil salinity, NDVI, EVI which are found to show a great impact on determining the yield. Without considering these features the accuracy is being compromised. Finally, this paper concludes that the consideration any one of features like wind, Velocity, soil texture, soil salinity, NDVI, EVI would increase the level of accuracy while predicting the yield and crop recommendation which gives profit to farmer.

Contribution of our paper:

Development of a novel approach to implementing the random forest algorithm for crop recommendation and yield prediction, tailored to the specific characteristics of the dataset and the study area.

Integration of multiple data sources, such as weather, soil, and crop management data, to improve the accuracy and precision of crop recommendation and yield prediction using the random forest algorithm.

Evaluation of the effectiveness and accuracy of the random forest algorithm for crop recommendation and yield prediction, and comparison with other commonly used algorithms, such as decision tree, naive bayes, and SVM.

Identification of the key variables and features that impact crop yield prediction using the random forest algorithm, and providing specific recommendations for farmers to optimize these variables to achieve higher crop yields.

VII. FUTURE ENHANCEMENT

The field of crop recommendation and yield prediction is constantly evolving, with many potential areas for future enhancement. These include the integration of additional data sources, such as crop health sensor data and social and economic data, as well as the development of personalized recommendations that take into account individual farmer characteristics and preferences. Advanced analytics techniques, such as deep learning and natural language processing, could be integrated into crop recommendation and yield prediction models to improve their accuracy and predictive power. Real-time monitoring and feedback systems could also be implemented, allowing farmers to make adjustments to crop management practices based on the predictions generated by the models. Finally, explainable AI methods could be incorporated to make these models more transparent and interpretable. Advancements in crop recommendation and yield prediction models have

https://doi.org/10.5281/zenodo.14190838

the potential to significantly improve the effectiveness and sustainability of agriculture.

VIII.REFERENCES

[1] Rajak, Rohit Kumar, et al. "Crop recommendation system to maximize crop yield using machine learning technique." *International Research Journal of Engineering and Technology* 4.12 (2017): 950-953.

[2]Y. Jeevan Nagendra kumar, et al. "Supervised machine learning approach for crop yield prediction in agriculture sector." *2020 5th International Conference on Communication and Electronics Systems (ICCES)*. IEEE, 2020.

[3] PANDE et al. "Crop recommender system using machine learning approach." *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 2021.

[4] Elavarasan, Dhivya, and PM Durairaj Vincent. "Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications." *IEEE access* 8 (2020): 86886-86901.

[5] Sonal Agarwal, et al. "Hybrid approach for crop yield prediction using machine learning and deep learning algorithms" 2021 *J.Phys.:Conf.ser*. 1714 012012.

[6] Bhatnagar, Kanika, Mamilla Jaahnavi, and BU Anu Barathi. "Agriculture Crop Recommendation System using Machine-Learning." *Mathematical Statistician and Engineering Applications* 71.3s2 (2022): 626-637.

[7] Priyadharshini, A., et al. "Intelligent Crop Recommendation System using Machine Learning." *2021 5th International Conference on Computing Methodologies and Communication (ICCMC)*. IEEE, 2021.

[8] Firdous Hina1, Dr. Mohd. Tahseenul Hasan et.al (2022) "Agriculture Crop yield Prediction Using machine learning" Anjuman College of engineering and technology.

[9] Swapneel Chakraborty et,al "Intelligent crop recommendation System using Machine Learning" BE Information science and engineering.

[10] M.A.Manivasagam et.al,"An efficient crop yield prediction using Machine learning",ISSN,2022.

[11] Meghana et al, "Crop and Yield Prediction with Fertilizer estimation" Dr. Ambedkar Institute of technology, Bangalore.

[12] Dhanush, Viswakarma, mahendra et al (2021) "Crop prediction using machine learning approaches" PES college of engineering, Department of E&C Engineering.

[13] Madhuri Shripathi Rao et al , Crop prediction using machine learning, Journal of Physics: Conference Series.

[14] Mansing Rathod, Chaitanya Chandvidkar, Darpan Chachpara, and Mayank Champaneri. prediction of crop yield using machine learning..

[15] *Kevin Tom Thomas, Varsha, Merin Mary Saji, Lisha Varghese, Er. Jinu Thomas "Crop Prediction Using Machine Learning" International Journal of Future Generation Communication and Networking Vol. 13, No. 3, (2020), pp. 1896–1901.*

[16] Jeevan Nagendra Kumar Y, V Spandana, VS Vaishnavi, K Neha, VGRR Devi. Supervised Machine learning Approach for Crop Yield Prediction in Agriculture Sector. 5th International Conference on Communication and Electronics Systems. 2020;736-741

[17] Nischitha K, Dhanush Vishwakarma, Mahendra N Worked on "Crop Prediction using Machine Learning Approaches" International Journal of Engineering Research & Vol. 9 Issue 08, August-2020.

[18]Janmejay Pant et al., Analysis of agricultural crop yield prediction using statistical techniques of machine learning,the scientific committee of the International Conference on Technological Advancements in Materials Science and Manufacturing.

[19] Zeel Doshi et al, AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms.

https://doi.org/10.5281/zenodo.14190838