

# CROP RECOMMENDATION ASSISTANCE USING MACHINE LEARNING (KNN ALGORITHM) AND PYTHON GUI

ADITYA GHOSH<sup>1</sup>, ANUBHAV SENAPATI<sup>1</sup>, RAHUL DAS<sup>1</sup>, SNEHA SARKAR<sup>1</sup>, JOYEE SAHA<sup>1</sup>, ANNWESHA MAHANTA<sup>1</sup>, SUDIPTA BASU PAL<sup>2\*</sup>

<sup>1</sup>Department of CST

University of Engineering and Management, Kolkata India

<sup>2</sup>Department of CST and CSIT

University of Engineering and Management, Kolkata  
India

Email: [ghosh.aditya111@gmail.com](mailto:ghosh.aditya111@gmail.com), [senapatianubhav1651@gmail.com](mailto:senapatianubhav1651@gmail.com), [physicsrahul8697@gmail.com](mailto:physicsrahul8697@gmail.com), [snehasarkar439@gmail.com](mailto:snehasarkar439@gmail.com), [joyeesaha2001@gmail.com](mailto:joyeesaha2001@gmail.com), [annweshamahanta0@gmail.com](mailto:annweshamahanta0@gmail.com), [sudipta\\_basu68@yahoo.com](mailto:sudipta_basu68@yahoo.com)

\*Corresponding Author: [sudipta\\_basu68@yahoo.com](mailto:sudipta_basu68@yahoo.com)

## Abstract-

Finding and extracting significant records from data is the undertaking of facts mining.. information mining has applications in various fields such as finance, retail, remedy, agriculture, etc. Agricultural records mining is used to research numerous biotic and abiotic elements. Agriculture in India plays a main position inside the economy and employment. A common trouble amongst Indian farmers is they do not choose the right crops in keeping with the desires of their soil. As a end result, they face extreme productivity setbacks. Precision agriculture solves this trouble for farmers. Precision farming is a modern-day agricultural technique that makes use of studies facts gathered on soil houses, soil types, crop yield information and recommends appropriate plants to farmers based on unique local parameters. This reduces poor crop choice and will increase productiveness. on this paper, this problem is solved by using offering a recommender machine that makes use of an ensemble version with random trees, CHAID, KNearest pals, and a majority voting technique with Naive Bayes as freshmen

**Keywords:** Machine Learning, KNN algorithm, PYTHON GUI, Data Mining

## I. INTRODUCTION

In India, agriculture is the main industry. Agriculture exports and imports have a significant impact on India's economy. One of the key sectors of the Indian economy is agriculture. The economic situation has drastically declined because of the uncertainty surrounding crop yield. Rice, wheat, pulses, and grains are India's primary agricultural products. India's population is expanding daily, necessitating improved crop

yield in order to feed the nation[1-5]. Use of machine learning algorithms is one of the greatest approaches to forecast unknown values. This project aims to create a machine learning-based crop prediction model. The program's goal is to forecast crop yield so that farmers may select the optimal seeds for planting. There are many machine learning (ML) algorithms that can be applied, including regression analysis, support vector machines, neural networks, and K-Nearest Neighbor (K-NN). In this paper, KNN is discussed. The k-nearest neighbors (KNN) algorithm is a straightforward, supervised machine learning technique that may be applied to both classification and regression issues[6-8]. Our goal is to employ a model in which information is concentrated in a few groups to forecast the categorization of a subsequent instance. K-NN calculates the k-nearest neighbors based on the shortest distance between the query instance and the training examples. The prediction query object is then chosen by a simple majority of the collected k nearest neighbors.

## II. LITERATURE SURVEY

In a research paper, Rashi Agarwal investigated machine learning methods. This method would assist farmers in selecting the best crops to plant based on a number of geographical and environmental parameters. They used neural networks, decision trees, KNNs, Random Forests, and more. The accuracy of the neural network was the highest. In her research piece, Priyadarshini A did a study on machine learning algorithms. By assisting farmers in selecting the right crop and providing the

data that conventional farmers do not maintain, technology reduces crop failure and lowers production. There were several different machine learning algorithms used. The accuracy leader among the group was the neural network. Shilpa Mangesh Pande She offers a practical and farmer-friendly production forecasting technique in her study article. A smart phone application connects the suggested technology to farmers. With the aid of GPS, the user's location can be ascertained. The accuracy of crop yield forecasts is compared across all algorithms. With a 95% accuracy rate, the RF algorithm proved to be the best for the given data set. A data mining technique was used in Mayank Champaneri's research on predicting crop yields. Because a random forest classifier can handle both classification and regression tasks, they used it. Anyone can use the userfriendly website that was created to anticipate crop yield for their preferred crop by providing climate data for that area.

### III. PROBLEM STATEMENT

In a country like India, many factors affect yield. Harvest predictions are highly dependent on variables such as humidity, temperature, rainfall and soil type, all of which vary greatly from place to place. Indian farmers continue to rely heavily on traditional techniques passed down from their ancestors. These techniques are more effective if the climate is more stable and predictable. Environmental issues such as global warming and pollution are impacting the environment, so people need to be smart and start using the latest technology now. It's time to go through the masses of information and develop a tool that can provide users with relevant information about crop productivity. New Age techniques require large structured data sets and algorithms that can generate solutions using the provided data.

### IV. PROPOSED SOLUTION

The novelty of the suggested approach is that it gives farmers instructions on how to increase crop yield while also recommending the most lucrative crop for a given area. Methodology of crop recommended systems are:-

**A. Dataset collection:** It might not be enough to merely take one or two elements into account when putting an accurate prediction model into practice. Data on temperature, humidity, rainfall, and other variables are gathered and examined.

**B. Data Preprocessing:** After collecting data from various sources, the next step is to preprocess the data before training the model. Starting with reading the captured data set and cleaning the data, data preprocessing can be done in ways. When cleaning Information some record characteristics are duplicated. As a result, you need to remove unnecessary properties and records containing missing data.

**C. Feature engineering:** It is the process of extracting features (characteristics, traits, and qualities) from raw data using domain expertise. The goal is to employ these additional attributes to raise the caliber of ML output.

**D. Training set:** A training set is a set of data that includes labelled data. Vectors for the input and output are both present. The model is trained using supervised machine learning algorithms using this dataset.

**E. Testing set:** A testing set is a data set that is devoid of labelled data. It predicts the outcome with the assistance of the training data set. It is unaffected by the training data set.

**F. K-Nearest Neighbors:** KNN is a supervised machine learning technique that can be applied to a variety of problems. Regression and classification are two instances of difficulties that can be addressed. The letter K stands for the quantity of closest neighbors to a newly predicted unknown variable. The distance between the data points is calculated using the Euclidean distance formula. Euclidean Distance b/w A and B =  $\sqrt{(X_2 - X_1)^2 + (Y_2 - Y_1)^2}$ .

**G. Crop Recommendation:** By examining variables like rainfall, temperature, area, humidity, contents of the soil, PH value etc., the suggested model forecasts crop yield.

**H. Performance Analysis:** It is a specialization that uses systematic goals to enhance performance and decision making.

### V. METHODS AND TECHNOLOGY MODULES

Text to Speech by using pyttsx3 pyttsx3 is a text-to-speech conversion library in Python. Unlike alternative libraries, it works offline and is compatible with both Python 2 & 3. An application invokes the pyttsx3.init() factory function to get a reference to a pyttsx3. Engine instance. it is a very easy to use tool which converts the entered text into speech. The pyttsx3 module supports two voices first is female and the second is male which is provided by —sapi5 for windows.

Pandas DataFrame is two-dimensional size-mutable, potentially heterogeneous tabular data structure with labeled axes (rows and columns). A Data frame is a two-dimensional data structure, i.e., data is aligned in a tabular fashion in rows and columns. Pandas DataFrame consists of three principal components, the data, rows, and columns[9-12].

Data Pre-Processing with Sklearn using Data Scaling is a data preprocessing step for numerical features.

Many machine learning algorithms like Gradient descent methods, KNN algorithm, linear and logistic regression, etc. require data scaling to produce good results. Various scalers are defined for this purpose.

This article concentrates on Standard Scaler and Min-Max scaler. The task here is to discuss what they mean and how they are implemented using in-built functions that come with this package.

Apart from supporting library functions other functions that will be used to achieve the functionality are: The fit(data) method is used to compute the mean and std dev for a given feature so that it can be used further for scaling. The transform(data) method is used to perform scaling using mean & std dev calculated

using the .fit() method. The fit\_transform() method does both fit and transform.

ML | Implementation of KNN classifier using Sklearn K-Nearest Neighbors is one of the most basic yet essential classification algorithms in Machine Learning. It belongs to the supervised learning domain and finds intense application in pattern recognition, data mining and intrusion detection. It is widely disposable in real-life scenarios since it is non-parametric, meaning, it does not make any underlying assumptions about the distribution of data (as opposed to other algorithms such as GMM, which assume a Gaussian distribution of the given data)[13-15].

Our model is based on K Nearest Neighbour (KNN) of Machine Learning. It is based on Supervised Learning technique. It uses 'feature similarity' to predict the values of new data points which further means that the new data point will be assigned a value based on how closely it matches the points in the training.

The flow chart is given in figure 1.

The KNN algorithm working can be explained on the basis of the below steps:

The flow chart is given in figure 1.

The KNN algorithm working can be explained on the basis of the below steps:

Step-1: Select the number K of the neighbours. Step-2:

Calculate the Euclidean distance of K number of neighbours.

Step-3: Take the K nearest neighbours as per the calculated Euclidean distance.

Step-4: Among these k neighbours, count the number of the data points in each category.

Step-5: Assign the new data points to that category for which the number of the neighbour is maximum.

Step-6: End of the algorithm.

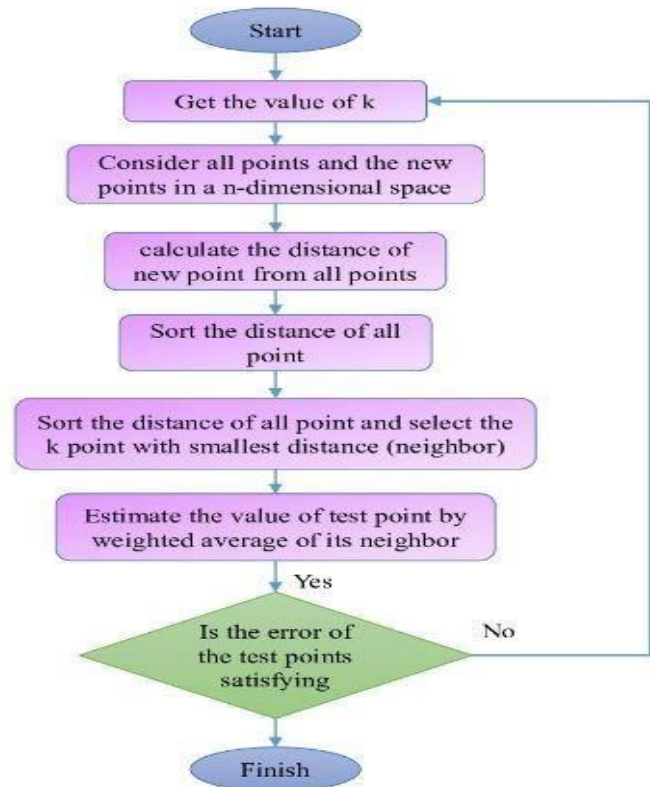


Fig1: Representation of the K nearest neighbours in flow chart diagram

#### A. Other Recommendations

PySimpleGUI : It is easy to use with simple yet HIGHLY customizable features of GUI for Python. It is based solely onTkinter. It is a Python GUI For Humans that Transforms Tkinter, PyQt, Remi, WxPython into portable user-friendly Pythonic interfaces.

#### VI. EXPERIMENTAL SETUP AND RESULT ANALYSIS:

To analyze that how KNN algorithm works and produces the desired outcome. We have compared it with K means algorithm in Table 1.

KNN ALGORITHM	K MEANS ALGORITHM
K-NN is a Supervised machine learning	K-means is an unsupervised machine learning.
KNN is classification or regression machine learning algorithm	It is a clustering machine learning algorithm.
Capable of Calculation of predicting error.	It can't do so.
Classes are already created	It creates classes
It makes predictions by learning from the past available data.	used for analyzing and grouping data

We have compared it with decision tree algorithm in Table2.

KNN ALGORITHM	DECISION TREE
Compare a new data points to similar labelled data points.	Use thresholds of feature values to determine classification
Easy to implement	Very hard to find globally-optimal trees
Best work when it comes to rare occurrences.	Take much time in that cases.

Table:2

**KNN vs. Linear Regression :**

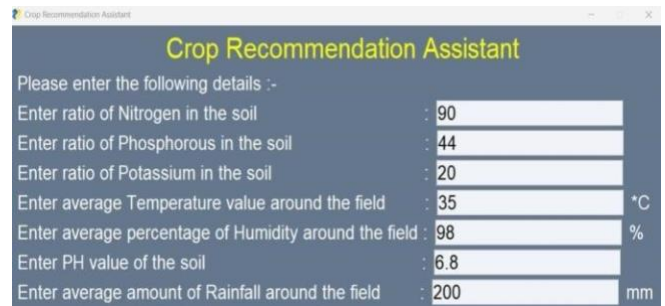
KNN is better than linear regression when the data have high SNR(Signal Noise Ratio).

**KNN vs. SVM (Support Vector Machine) :**

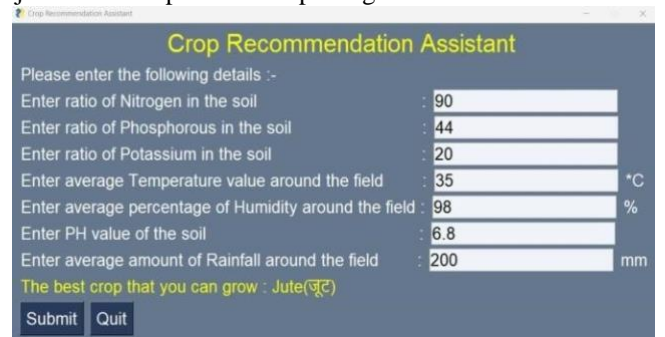
If training data is much larger than no. of features(m>n), KNN is better than SVM.

**KNN vs. Naive Bayes :**

KNN doesn't require any training—you just load the dataset and off it runs. On the other hand, Naive Bayes does require training. Now analyzing the result, The Inputs are given as follows in figure 2—



With the given inputs, our Crop Recommendation Assistant projected Jute as potential crop in figure 3.



THIS ASSISTANT LOWERS UNNECESSARY COSTS AND AIDS IN ELIMINATING THE USE OF SENSORS. TIME AND MONEY ARE USED EFFICIENTLY BY THIS SYSTEM.OUR SYSTEM ASSISTS IN GATHERING ALL RELEVANT DATA AND PROVIDING AN OUTPUT MODEL THAT NOT ONLY BOOSTS PRESENT ECONOMIC GAIN BUT ALSO ENSURES FUTURE PROFITABILITY.ALTHOUGH CONSIDERED TO BE DECENT, THE SYSTEM'S ACCURACY COMPONENT MIGHT BE IMPROVED WITH GREATER EFFICIENCY.

**VII. CONCLUSION**

It has been seen that a lot of agricultural research has been done and is still being continued to increase production, strengthen the Indian economy, and, most significantly, help farmers earn more money. The proposed system will inform farmers of the optimal crop to grow on their selected land in order to achieve this. The system was put into place so that people, who are new in this field, could learn about farming and crops and discover effective harvesting techniques. The study primarily uses agricultural records from numerous portals that belong to a few districts. For the prediction model and crop yield prediction, the K-NN algorithm is employed, and its accuracy is reached. The application of machine learning algorithms in crop production has a promising future as we intend to use more sophisticated algorithms to make the system more effective. We also hope to use more datasets and sophisticated algorithms to make system prediction more stable and achieve high.



## VIII. FUTURE SCOPE

Our future task is to improve the results of this model i.e. we want to add few features such as yield prediction, price forecasting, plant growth assistant etc.using large number of crop type datasets, soil type datasets and more weather parameters. Building a strong technical partner/assistant model which will predict yield and forecast price and will provide a suitable alarm for all the necessary task such as watering the plant, manure time, etc. for all the crops based on the analysis. Generating the voice part of crop recommendation using natural or local language to make it user friendly.

## REFERENCES

- [1] Tanmay Banavlikar et al. —Crop recommendation system using Neural Networks. In: International Research Journal of Engineering and Technology (IRJET) 5.5 (2018), pp. 1475–1480.
- [2] Pradeepa Bandara et al. —Crop recommendation system. In: International Journal of Computer Applications 975 (2020), p. 8887.
- [3] Gouravmoy Banerjee, Uditendu Sarkar, and Indrajit Ghosh. —A Fuzzy Logic-Based Crop Recommendation System. In: Proceedings of International Conference on Frontiers in Computing and Systems. Springer, 2021, pp. 57–69.
- [4] Zeel Doshi et al. —AgroConsultant: intelligent crop recommendation system using machine learning algorithms. In: 2018 Fourth International Conference on Computing Communication Control and Automation (ICCC3). IEEE, 2018, pp. 1–6.
- [5] Dhruvi Gosai et al. Crop Recommendation System using Machine Learning. 2021.
- [6] Nidhi H Kulkarni et al. —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). IEEE, 2018, pp. 114–119.
- [7] Javier Lacasta et al. —Agricultural recommendation system for crop protection. In: Computers and Electronics in Agriculture 152 (2018), pp. 82–89.
- [8] Daneshwari Modi et al. —Crop Recommendation Using Machine Learning Algorithm. In: 2021 5th International Conference on Information Systems and Computer Networks (ISCON). IEEE, 2021, pp. 1–5.
- [9] A Priyadarshini et al. —Intelligent Crop Recommendation System using Machine Learning. In: 2021 5th International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2021, pp. 843–848.
- [10] S Pudumalar et al. —Crop recommendation system for precision agriculture. In: 2016 Eighth International Conference on Advanced Computing (ICoAC). IEEE, 2017, pp. 32–36.
- [11] Rohit Kumar Rajak et al. —Crop recommendation system to maximize crop yield using machine learning technique. In: International Research Journal of Engineering and Technology 4.12 (2017), pp. 950–953.
- [12] D Anantha Reddy, Bhagyashri Dadore, and Aarti Watekar. —Crop recommendation system to maximize crop yield in ramtek region using Machine learning. In: International Journal of Scientific Research in Science and Technology 6.1 (2019), pp. 485–489.
- [13] Rohit Kumar Rajak et al. —Crop recommendation system to maximize crop yield using machine learning technique. In: International Research Journal of Engineering and Technology 4.12 (2017), pp. 950–953.
- [14] D Anantha Reddy, Bhagyashri Dadore, and Aarti Watekar. —Crop recommendation system to maximize crop yield in ramtek region using Machine learning. In: International Journal of Scientific Research in Science and Technology 6.1 (2019), pp. 485–489.
- [15] M Rekha Sundari et al. —Crop Recommendation System Using K-Nearest Neighbors Algorithm. In: Proceedings of 6th International Conference on Recent Trends in Computing. Springer, 2021, pp. 581–589.