CROP RECOMMENDATION SYSTEM USING KNN ALGORITHM AND RANDOM FOREST

V.Nagesh[1], M. Moulika[2], K. Vaishnavi[3], K. Bhavishya[4], M. Nikitha[5].
[1]Associate Professor, [2,3,4,5]Sreyas Institute of Engineering and Technology

Abstract:
In agriculture, the integration of machine learning has been a long-standing aspiration, resulting in significant advancements. While machine learning models have been developed for crop and yield predictions, traditional algorithms like decision trees often fall short of delivering the desired accuracy. This paper introduces an accessible and user-friendly solution for crop recommendations and yield predictions. Users provide inputs such as temperature, humidity, soil pH, and rainfall. To enhance accuracy, a hybrid approach using K-nearest neighbor (KNN) and Random Forest (RF) algorithms is employed. The K-nearest neighbor (KNN) algorithm achieves an impressive accuracy rate of 98%. Additionally, the Random Forest (RF) algorithm attains a commendable 96% accuracy by aggregating multiple decision trees. These high accuracy rates signify the system's potential to empower farmers with data-driven insights for crop selection and yield projections. Furthermore, the user-friendly interface promises broader adoption within the agriculture sector, catering to users with varying levels of technical proficiency. To strengthen the system's credibility, transparency regarding data sources and quality is imperative. Utilizing accurate and relevant data for reliable predictions. In summary, this paper presents a promising solution for informed decision-making in agriculture, combining crop recommendations and yield predictions. Acknowledging the limitations of traditional approaches, it capitalizes on the strengths of K-nearest neighbor and Random Forest algorithms.

Keywords: Crop recommendation, Yield prediction, Machine learning, KNN, Random Forest

1 Introduction:

Agriculture, as one of humanity's oldest and most vital practices, has continually evolved to meet the continuously increasing global demand for food, fibre, and biofuel production. However, agriculture faces a host of complex challenges, ranging from climate change-induced variability in weather patterns to natural resources and the importance of sustainable land management. In response to these challenges, the fusion of modern technology and machine learning has given rise to Crop Recommendation Systems (CRS), a pioneering approach that promises to redefine the way farmers make crucial decisions about crop selection and cultivation practices. The traditional approach to farming often relies on generational knowledge, local practices, and some basic observations. Yet, this method frequently results in suboptimal yields, resource misallocation, and vulnerability to the unpredictability of weather patterns. Crop Recommendation System, powered by machine learning algorithms, represents an important change in agriculture, offering data driven, precision farming solutions that can significantly enhance productivity, profitability, and sustainability. This paper embarks on a journey into the realm of Crop Recommendation Systems, inquiry into their importance, the technological foundation that makes them feasible, and their real-world implications. By the end of this exploration, readers will gain a comprehensive understanding of how machine learning is composed to revolutionize agricultural decision-making and contribute to a more sustainable, productive, and food-secure world.

Crop Recommendation System marks a departure from general farming approaches towards precision agriculture. These systems force historical and real-time data, including soil properties, climate conditions, and past crop performances, to offer farmers tailored recommendations that optimize resource allocation, reduce waste, and maximize yields. In an era of existing concern over environmental sustainability, the Crop Recommendation System plays a crucial role in promoting eco-friendly agriculture. By suggesting crop varieties resilient to changing climate conditions and advocating for sustainable land management practices, these systems contribute to long-term environmental preservation. With the global population continuously on the rise, ensuring food security is a pressing global challenge. Crop Recommendation System has the potential to support food production by guiding farmers toward more productive crop choices, helping meet the growing demand for food even amidst uncertain climatic conditions.Crop Recommendation Systems are already making a substantial impact in agriculture globally. For instance. This mobile application utilizes real-time weather data and crop-specific information to provide Indian farmers with guidance on crop selection, disease management, and weather forecasts. It has significantly improved the decision-making abilities of millions of small-scale farmers. IBM's platform deploys AI and machine learning to offer insights into soil health, crop diseases, and pest infestations. It assists in optimizing planting, harvesting, and irrigation schedules. As these technologies continue to evolve and adapt to diverse agricultural landscapes, they hold the promise of uplifting the livelihoods of farmers worldwide while ensuring a sustainable and efficient future for global agriculture. This paper will further explore the mechanics and applications of these systems, providing an indepth overview of their pivotal role in shaping the future of agriculture.

The proposed system focuses on following major contributions. 1. Crop Recommendation Systems, driven by machine learning, stand at the forefront of agricultural innovation. They harness the potential of data-driven decision-making to usher in an era of higher agricultural yields, reduced environmental impact, and enhanced food security.

2. Here we state the objectives of the system, such as improving crop yield, resource optimization, and providing farmers with actionable insights.
3. The sources and types of data to be collected such as weather data, soil data, farmer-specific data. We provide data preprocessing steps, such as data cleaning and feature extraction. In detail then we depict the comprehensive crop database, including crop characteristics, growth requirements.
4. Here we make use of two machine learning algorithms: K-Nearest Neighbors (KNN) and Random Forest. The suitability of KNN for similarity-based recommendations and Random Forest.
5. We emphasize the importance of real-time and personalized recommendations using the technology stack and tools in system development and provide a high-level system architecture diagram.

6. We provide the integration of real-time monitoring systems for weather updates, soil moisture, and pest/disease alerts, analytics and insights provided to farmers, including yield predictions, profit margins, and resource usage statistics.
7. Discuss the mechanism for collecting feedback from farmers and how it will be used to improve the system and detail the development of a mobile application to ensure accessibility for remote farmers and the ongoing maintenance and support plan.
8. This proposed system combines the power of KNN and Random Forest algorithms to provide accurate and robust crop recommendations, ensuring all the components and objectives of the system are well-documented and considered during development.

2 Literature Survey:

Rajak et al. [1] in the proposed work of crop recommendation system considers the soil dataset as for four harvests rice, cotton, sugarcane, and wheat. The dataset is first pre-processed and afterward the group method plays out a basic capacity in the characterization of the four yields. The individual base learners utilized in the Crop Recommendation System Using K-Nearest Neighbors Algorithm 583 ensemble model are random forest, Naive Bayes, and linear support vector machine (SVM). Greater part voting technique has been utilized as the blend strategy to give the best exactness. But the drawback is that it does not check the present soil condition and directly assume the soil type from past data and recommend the crop.

Pudumalar et al. [2] in their work about exactness horticulture, which is an advanced cultivating procedure utilizes research information of soil attributes, soil types, crop yield information, and recommends the ranchers the correct harvest dependent on their site explicit boundaries. This diminishes an inappropriate decision on a yield and expands efficiency. In this paper, this issue is settled by proposing a suggestion framework through an ensemble model with majority voting technique using random forest, k-nearest neighbor and Naive Bayes as learning algorithms to suggest a harvest for the site explicit boundaries with high exactness and proficiency. Kumar et al. [3] states that the farmers are commonly not mindful of the necessities of the yield, for example, the minerals, soil dampness, and other soil necessities. This can cause pain to farmer both psychologically and economically. One more issue that a farmer may largely experience is absence of information about vermin and sicknesses that can influence the harvests they develop. This issue of farmer is addressed in this paper and the authors attempted to understand it with the assistance of a Recommendation System. By the assistance of this model, they anticipated the best reasonable harvest to the rancher, recognized the bug that may influence and proposed the nuisance control strategies. In this paper, they have applied SVM classification algorithm, decision tree calculation, and logistic regression calculation and found that SVM model gives the better precision when contrasted with different calculations.

Gandge et al. [4] in their work used datasets with various variables like soil quality, nitrogen (N), potassium (K), phosphorous (P), and moisture contents over a period of time. They worked on different algorithms on different crops.
Multiple linear regression on rice crop. KNN algorithm on wheat and potato crops. Neural networks on corn yield. But there was no algorithm which suits for any crop from this project.

Yan et al. [5] considered the relationship between wheat growth and meteorology as their research object. Basically, thirteen features like atmospheric pressure, temperature, light, and precipitation from surface meteorological data are considered, but after discretization, they have selected only three features accumulative temperature, sunshine hours, and temperature. Using Apriori, a model was constructed for selecting farming season with the month as index.

Priya et al. [6] in the proposed work analysed the data according to seasons. Dataset contain factors like rainfall, soil moisture, temperature, and atmospheric pressure. Map reduce and Naïve bayes classifier were used to predict the crop. The result was depicted and differentiated according to seasons for different crops. This paper recommends us when to sow but not what to sow to get more yield production with specific dataset factors.

Pawar and Chillarge et al. [7] in their work used J48 algorithm for classifying fertility and predict the soil toxicity of soil. Features like pH value and nutritional contents (like P, N, Cl, Ca). Accuracy of sensors plays crucial role in this project. It tests the soil and recommends fertilizers to be added to increase the crop production, water level. It does not recommend crop which gives profit 584 M. Rekha Sundari et al. for farmers by yielding more production but it predicts the toxicity of soil and its fertility.

Kumar et al. [8] in their proposed work considered features like rainfall, pH value, and temperature. SVM, decision tree, and random forest algorithms were used but the accuracies of algorithms were 89.66, 86.8, and 88%, respectively, which did not cross at least 90%. This system was designed using Spyder IDE.

Priya et al. and Priyanka et al. [9,10] only depicted how algorithms can be used in recommendation of the crop but has not validated the efficiency of algorithms in performing the task.

Sl.No	Year of publication	Author	Results/comments
1	2017	Rajak et al. [1]	CRS to maximize crop yield using machine learning technique.
2	2017	Pudumalar et al. [2]	CRS for precision agriculture.
3	2018	Yan et al. [5]	Research on precision management for farming season based on big data.
4	2018	Priya et al. and Priyanka et al. [9,10]	Agricultural crop yield using machine learning and artificial intelligence and satellite imagery.

5	2019	Kumar et al. [8]	Recommendation system for crop identification and pest control techniques.

3 Proposed System:

The task crop recommendation system is of greater importance in the field of agriculture. Therefore, developing a crop recommendation system has gained much interest in the field of research nowadays. The task of this crop recommendation system is to recommend crop type and yield of the crop for various states of India based on a variety of factors like nitrogen level of the soil, phosphorous level of the soil, potassium level of the soil, season, and district. The current paper focus on Machine learning which includes supervised learning models like KNN and Random Forest. To recommend the crop a K- Nearest Neighbour algorithm is used, the output from the KNN is fed to the Random Forest which will provide us the desired outcome. System architecture is one which defines the Conceptual model in different structures and multiple views of the system. Fig. 1 shows architecture design of proposed system of the project.

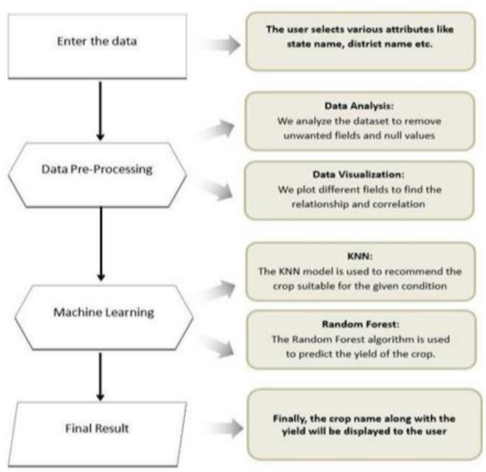


Fig. 1. Process flow Architecture

Above architecture clearly explains about how the components of the system communicate among themselves starting from data entry by user. This proposed framework is able to finding out the crop and its yield. This model gives clear picture of huge amount of data capture and preprocessing of data to remove the unwanted data such as NULL etc presented in it. During preprocessing step, we split the dataset into training and testing dataset. Train dataset to detect the crop yield present in the dataset using appropriate supervised learning algorithms. Apply the machine learning techniques which are helpful for finding crop and its yield for any new data occurred in the data. After this data acquisition suitable machine learning algorithm must be applied to compute efficiency and capability of the model, here we have applied various machine learning algorithms like KNN and Random Forest.

Prediction of crop is done using KNN classification. The system can compute certain crops by using the value of humidity, temperature and soil properties. By using this system farmers will be able to increase agriculture productivity. Hence, this will have an overwhelming impact on poverty alleviation, boosting employment rate, human resource development and food security. The data set that we used to train our system was collected from the official website of Indian Agricultural Statistics and Indian Remote Sensing.

- It is a user-friendly interface, and it can be into a decision transparency.
- Helps in increasing the cost accounting problems and time saving problem.
- It helps in improving scalability and accessibility.
- Enhances the crop yield.

Packages that we have used while developing the ML modelare:

Scikit: Used for data pre-processing

Pandas: For performing certain operations on data

Matplotlib: For Data Visualization part ***Dataset:***

This machine learning model is built using benchmark dataset. We have collected our benchmark dataset from a wellreputed website called Kaggle. We have collected two data set one for recommending the crop and one for predicting the crop.

Attributes or variables in dataset are, a) State name

b) District name

c) Season

d) Crop

e) Area

f) Production

g) Nitrogen level

h) Phosphorous level

i) Potassium level

j) Rainfall

k) Humidity

l) Temperature

3.1 Data Preprocessing:

Data preprocessing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. Data preprocessing involves data analysis and data visualisation..

3.2 Data Analysis:

In Data analysis, the main focus falls on removing unnecessary fields and null values from the dataset in order to infer meaningful insights from the results thereby improving the efficiency of the model.
This helps make sure that data is evenly distributed, and the ordering does not affect the learning process. Basically, there are two types of data analysis techniques they are Qualitative analysis and Quantitative analysis.

3.3 Data Visualisation:

Data visualisation is the process of visualising the attributes in a given dataset. Data visualisation techniques like scatter plots and heat maps help us to identify which attributes have the best effect on the dependent variable.

3.4. K-Nearest Neighbor (KNN):

The K-Nearest Neighbor algorithm is a classification algorithm that takes a bunch of labeled points and uses them to learn how to label other points. K-NN is a non-parametric algorithm, which means it does not make any unnecessary assumption on underlying data. This algorithm is also called lazy learner algorithm. KNN paradigm: Similar cases with the same class labels are near each other. Thus, the distance between two cases is the measure of dissimilarity.

KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data.

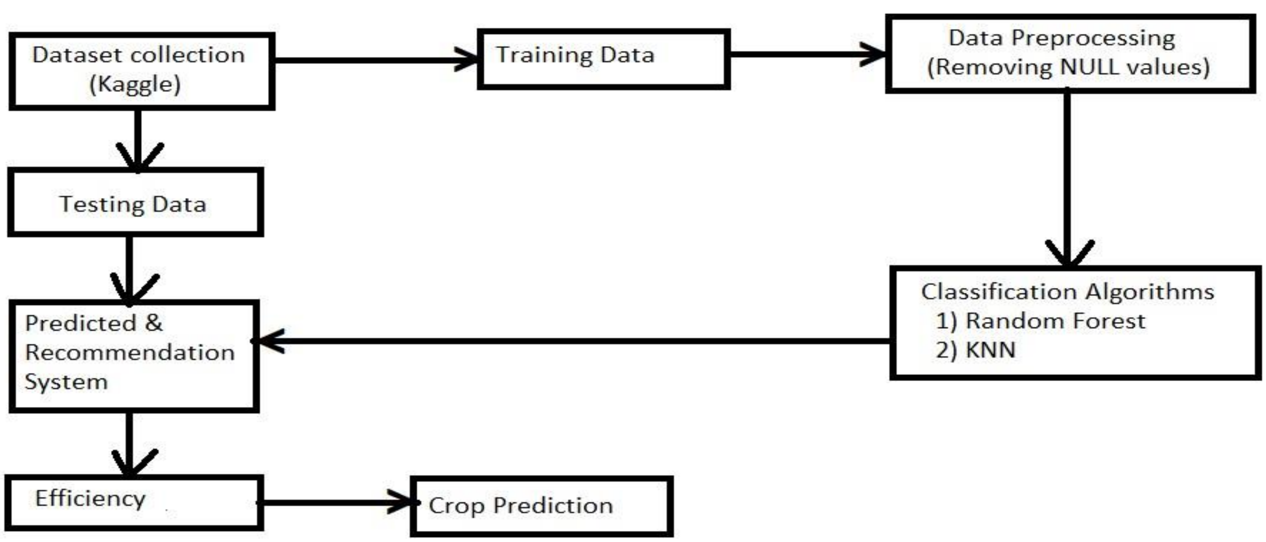
In our project the KNN machine learning model is used for recommending crop for area based on the input given by user such as potassium level, phosphorous level, rainfall, soil temperature, pH value of the soil, crop data, season and climatic parameters. Then the extracted data is fed to the machine learning model that is KNN. The KNN model performs certain analysis and gives output. The result is class membership. A majority vote of its neighbors classifies an object, with the object assigned to the class most common among its k nearest neighbors. If k=1, the object is simply assigned to the class of the object's single nearest neighbor

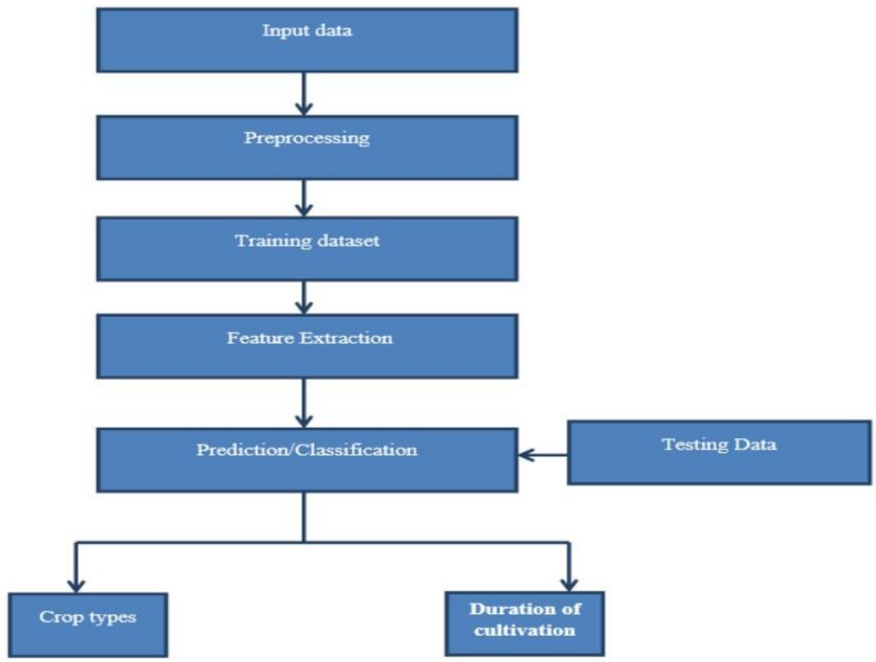
3.5 Random Forest:

Random Forest is a machine learning algorithm that can be highly effective in developing crop recommendation systems. It is based on the concept of ensemble learning, which is the process of combining multiple classifiers to solve a complex problem and improve the model's performance.

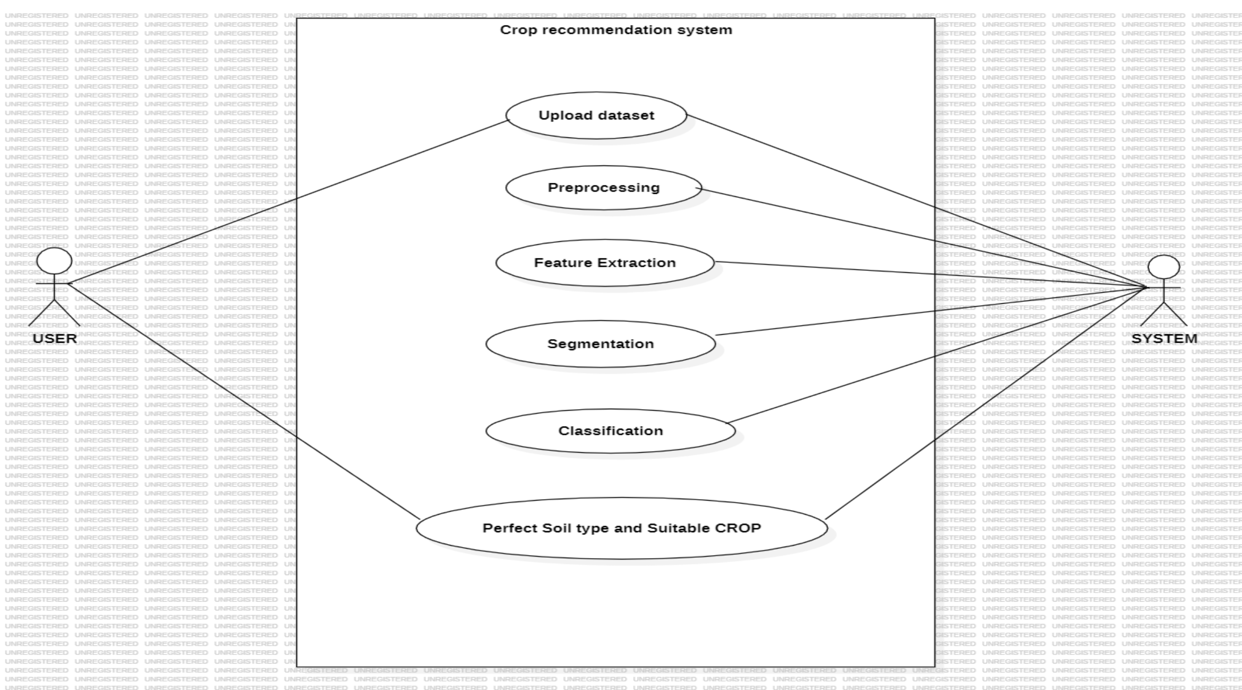
This classifier uses the average of a number of decision trees from a given dataset to improve the predictive accuracy of the dataset instead of depending on a single decision tree, the random forest takes the predictions from each tree and predicts the final output based on the majority vote of predictions made.

4 System Architecture:

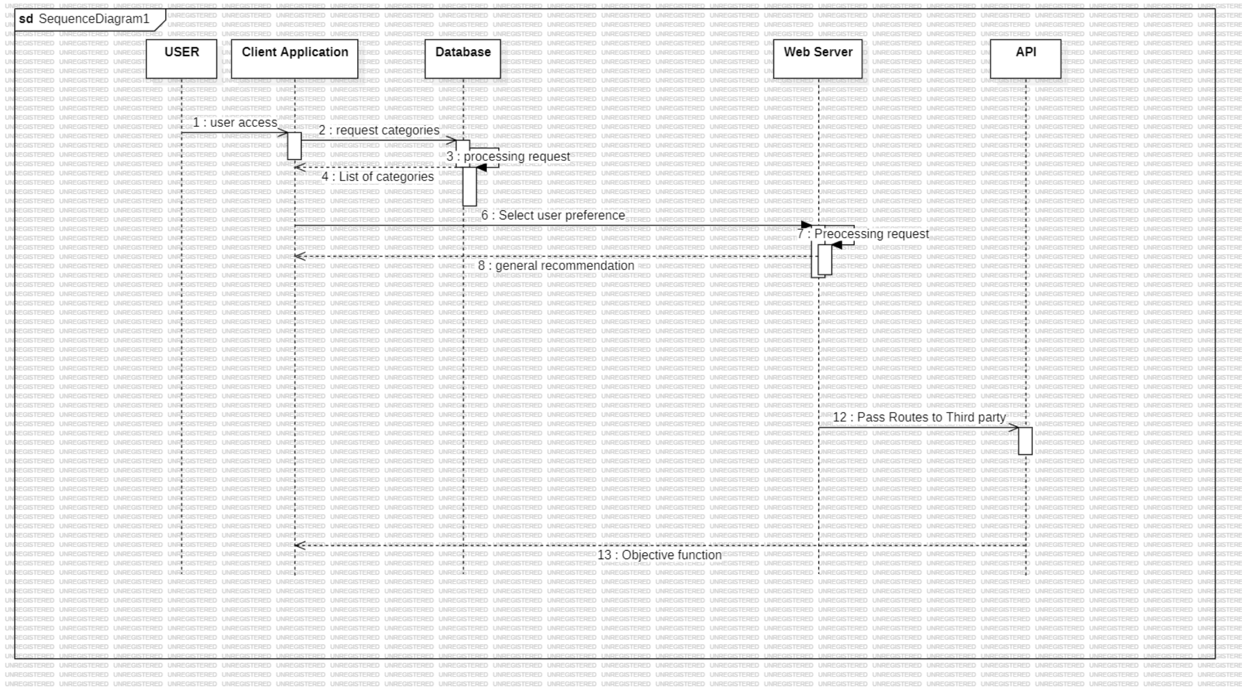




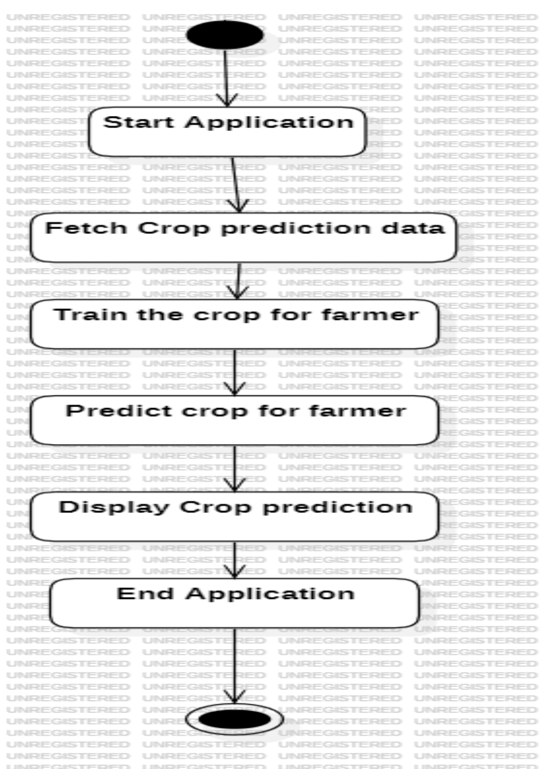
UML DIAGRAM



SEQUENCE DIAGRAM



ACTIVITY DIAGRAM



5 Result:

The collected data is initially subjected to pre-processing and then architectural flow diagram has been shown in figure 2. Post dataset pre-processing, the dataset is divided into training set and test set samples. Out of the 100% samples, 75% samples are used as training samples, and the rest 25% samples are used as test samples. Each of the samples is trained and tested on the Random Forest algorithm. The average accuracy of crop recommendation is 91.99%.

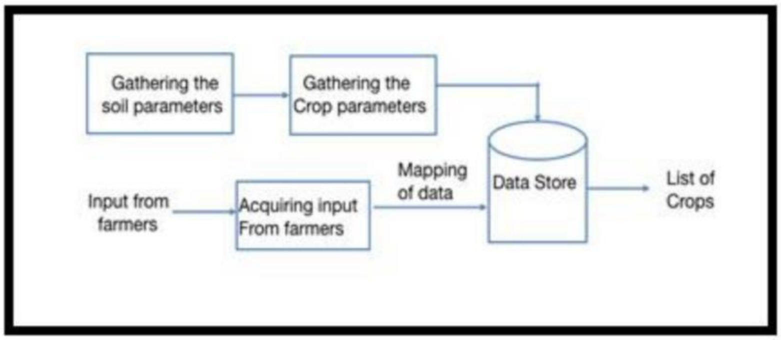


Figure 2 : Architectural diagram for crop recommendation system

	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
...	...	...	...	...	...	...	...
246086	West Bengal	PURULIA	2014	Summer	Rice	306.0	801.0
246087	West Bengal	PURULIA	2014	Summer	Sesamum	627.0	463.0
246088	West Bengal	PURULIA	2014	Whole Year	Sugarcane	324.0	16250.0
246089	West Bengal	PURULIA	2014	Winter	Rice	279151.0	597899.0
246090	West Bengal	PURULIA	2014	Winter	Sesamum	175.0	88.0

Table 1. Attributes for data analysis

The above table shows all the fields that are present in the dataset. It contains attributes such as state name, district name, crop year, season, area, production. The total number of rows present in the dataset are 246090 rows and total number of columns present in the dataset are 7 columns.

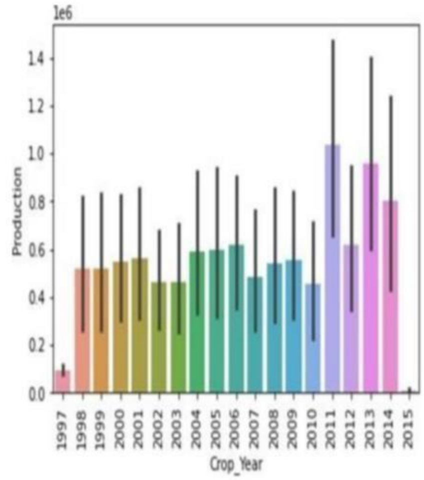


Fig 3. Data visualization for crop production in year

The above figure represents the Data visualization for crop production in year which is used for identifying dependent variable in the dataset.

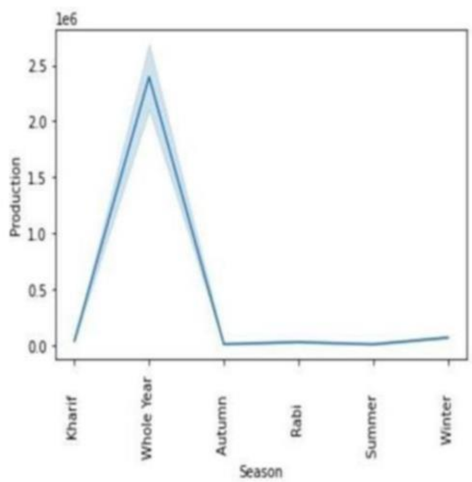


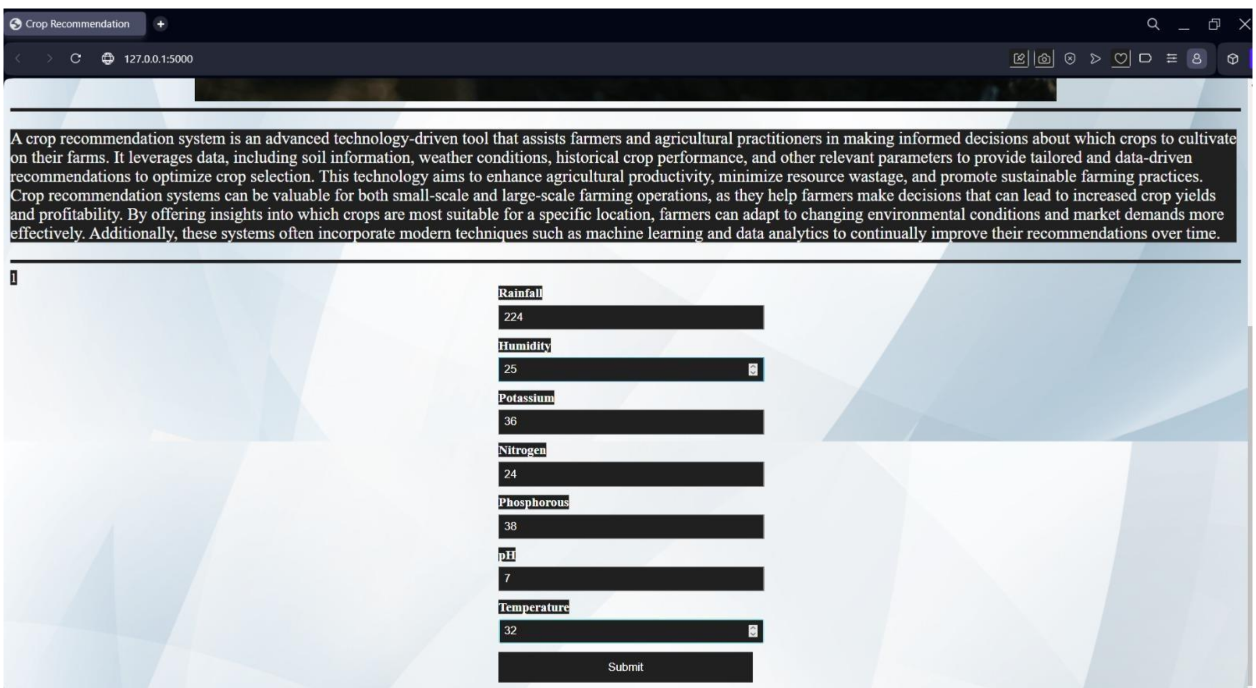
Fig 4. Data visualization for crop production in season

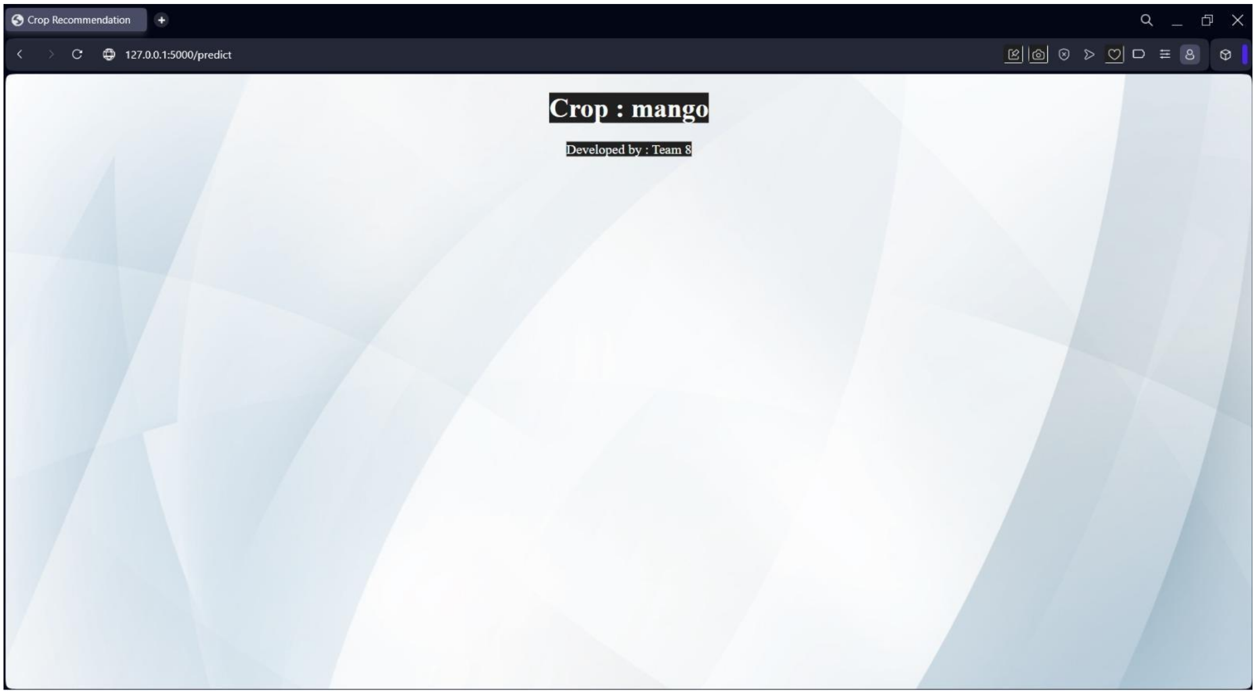
The above figure represents the Data visualization for crop production in season which is used for identifying dependent variable in the dataset.

As stated in the methodology, the historical data of crops grown under various soil parameters are taken into consideration while predicting the best crop that is to be grown in the present soil parameters. KNN classifier, the simple and easy to interpret classification algorithm is applied on various percentages of training data and test data with different values of k to check the accuracy. The algorithm gave an accuracy of 98.1% at K = 5 for training data.

The effectiveness of a crop recommendation system employing both the KNN algorithm and the random classifier in aiding farmers. KNN, known for its classification based on similar traits, complements the random classifier's diverse recommendations via random selection. Their amalgamation aims to mitigate biases and enhance accuracy, potentially curbing crop losses and bolstering farmers' profits. This system's value lies in its potential to refine agricultural practices, fostering sustainable farming. Continual exploration and enhancement of these algorithms are crucial for refining precision and effectiveness, underscoring the significance of ongoing research to optimize this agricultural tool.

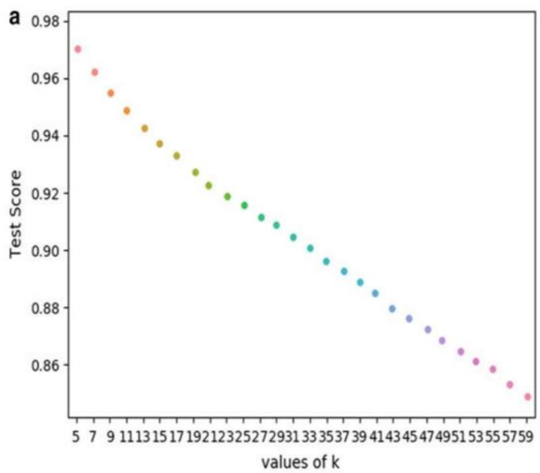
HOME PAGE

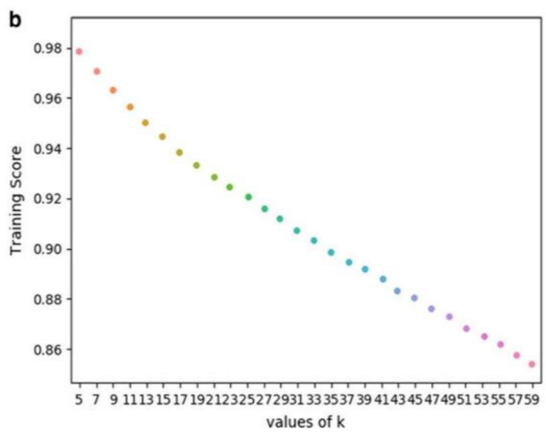




6 Conclusion:

In conclusion, the crop recommendation system using the KNN algorithm and random classifier is an effective tool for providing accurate recommendations to farmers. The KNN algorithm is able to classify crops based on similar characteristics and patterns while the random classifier uses a random selection method to provide diverse recommendations. The combination of these two algorithms helps to reduce bias and increase the accuracy of crop recommendations. This system can help farmers reduce crop losses and increase their profits. Overall, the crop recommendation system using the KNN algorithm and random classifier is a valuable tool for improving agricultural practices and promoting sustainable farming. It is essential to continue exploring and refining these algorithms to improve the precision and effectiveness of this system. In the proposed work, KNN algorithm is implemented on the data collected from different sources. As our future work, we aim to collect the features of the soil by using IOT devices with sensors that measure moisture and temperature so that we can predict the crop according to the changing features of the crop instantly without facing the difficulty of taking the samples to the lab. As KNN is lazy learning algorithm, we implement the methodology with eager learning algorithms by including other features like electrical conductivity and texture of soil to improve the predictive accuracy of the crop to be yielded.





References:

[1].Rajak RK et al (2019) Crop recommendation system to maximize crop yield using machine learning technique.
Int Res J Eng Technol 4(12):950–953

[2].Pudumalar S et al (2019) Crop recommendation system for precision agriculture. In: 2016 eighth international conference on advanced computing (ICoAC). IEEE

[3].Kumar A, Sarkar S, Pradhan C (2019)Recommendation system for crop identification and Pest control technique in agriculture. In: 2019 international conference on communication and signal processing (ICCSP). IEEE

[4].Gandge Y (2019) A study on various data mining techniques for crop yield prediction. In: 2017 international conference on electrical, electronics, communication, computer, and optimization techniques (ICEECCOT). IEEE

[5].Yan M et al (2018) Research on precision management of farming season based on big data EURASIP J Wireless Commune Networking 2018(1):143

[6].Priya R, Ramesh D, Khosla E (2018) Crop prediction on the region belts of India: a Naïve Bayes MapReduce precision agricultural model. In: 2018 international conference on advances in computing, communications and informatics (ICACCI). IEEE

[7].Pawar M, Chillarge G (2018) Soil toxicity prediction and recommendation system using datamining in precision agriculture. In: 2018 3rd international conference for convergence in technology (I2CT). IEEE

[8].Priya P, Muthaiah U, Balamurugan M (2018) Predicting yield of the crop using machine learning algorithm. Int J Eng Sci Res Technol 7(1):1–7

[9].Priyanka T, Soni P, Malathy C (2018) Agricultural crop yield prediction using artificial intelligence and satellite imagery. Eurasian J Anal Chem 13(SP):6–12. ISSN: 1306-3057