Farm-assist: Comprehensive Approach to Assist Farmers in the Agricultural Domain

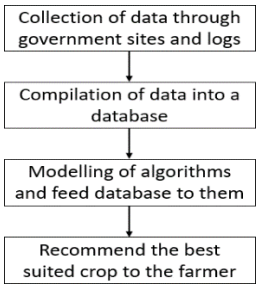
Bhawna Menghani, Bhawna Jain, Ayush Maithani, Jatin Garg
Department of Computer Science Engineering, Meerut Institute of Engineering and Technology, Meerut, India
Email: bhawnamenghani143@gmail.com, bhanuj2002@gmail.com, ayush.maithani07@gmail.com, sonamrahul96@gmail.com

Abstract—India's economy and employment are heavily reliant on agriculture. The most common problem faced by farmers of India today is their inability to select an ideal crop that depends upon the characteristics of region and past yields where they live. It therefore faces a large drop in terms of productivity. The administration has not explored agricultural statistics and projections to the extent necessary to fully understand their significance. The article introduces a portable, intelligent tool that uses machine learning techniques to assist farmers in selecting the best crops based on local climatic and soil conditions as well as other geographical characteristics. The sources available to farmers to answer all of their questions on seeds, fertilisers, market circumstances, storage facilities, government programmes, etc. are insufficient. A chatbot that employs the Deep Learning approach gives the farmers relatively easy access to this data analysis. In order to facilitate farmer contact, it is helpful to provide response to the input queries addressing the agricultural setting in audio format. If the system is unable to respond in any way, the inquiries are forwarded to helpline centres. The system's main function is to serve as a convenient and virtual helper for farmers all year long. It enables them to remain vigilant about every element that can affect agricultural productivity and revenue. Several machine learning algorithms that are built around the data set are used to generate the responses. Even though the system's primary goal is to support farmers more, anyone else who uses it, including children, can utilise it to gain help with all tasks associated with agriculture.

Keywords—Machine Learning, Deep Learning, K-Nearest Neighbour (KNN), Crop suggestion, Chatbot, Assistant, Agricultural Statistics, Application Programming Interface.

Introduction — The primary industry for employment is agriculture in many regions of the world. However, because they reside in rural areas, many farmers continue to be less informed. Field officers often interact with farmers in the communities and visit the fields to provide training and advice on the most appropriate farming techniques and farming elements [2][4]. In the present state of affairs, the government is collecting data in its raw form about rainfall and crop production but this data is not of any use to the customer, mainly who are farmers. It can be useful to the farmers if and only if the collection of raw data is standardized, analyzed, and then fed to the system that will provide the relational trends to them. Day by day, technology in the field of farming is expanding. Also, at the same time to educate the farmers about this technology, a large number of software are developed too. Most software provide stable farming information, require a large number of search steps and keen observation to get accurate facts, and do not provide an intercommunicating cross-questioning method. The proposed method gets over the aforementioned drawbacks by communicating properly with smoother transitions to get the desired result. [1][2]. This system, Farm-assist, is a virtual assistant that gives farmers a very practical and transportable instrument that might perform both of these tasks, namely, help them with farming techniques and recommend the best crop according to the given conditions, which will help them. in deciding what sort of crops can be grown to maximize yield in an efficient way. An application is designed and developed that includes a user interface for the farmers from certain places . The UI includes many features like weather forecast, recommendations on crops, doubt fields and Kisan help centres. It also provides a special local

language selection feature that allows local people who are very less exposed to mobile chatting and technology to interact with the chatbot.



View of the system in Fig. 1.

Beginning with collecting data, compiling data, feeding the data into a database and modeling that datawith different kinds of algorithms to recommend the best crops to farmers, we will discuss the flow of data through the system.

The system flow design for the entire project is depicted in Figure 2. We outline all the different types of parameters that are taken into account when the system is programmed to recommend the best crops.Then, data is processed using various machine learning techniques to train the system.

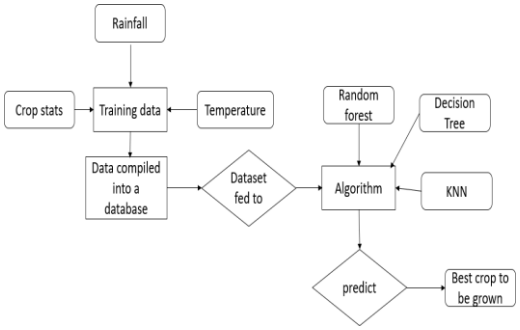


Fig. 2. System design flow chart

I. LITERATURE SURVEY

In this area, numerous analysis investigations are carried out. According to the rule selection methodology, they are distinguished from one another. For the project, the following literature was documented:

- In [1], the problem of Indian farmers' optimal crop selection for soil quality is addressed. Recommend systems have been developed by using various classification algorithms [11]. Hence, the system is operated via a graphical user interface.
- Exactitude farming's evolution and modelling are discussed in [2]. This system's primary goal is to directly inform farmers via mails or emails about various advising initiatives [12].
- In order to determine which algorithm is best for estimating the yield of different crops in Precision Agriculture, numerous algorithms are compared in [3]. The main dataset used to test each method is a multi-year set of soy bean cultivated data.The comparison algorithm used here includes Random Forest, SVM(support vector machine), Decision Tree with various other algorithms [13].
- Farm-assist is a chatbot in [4], i.e., a virtually designed assistant that makes it simple for consumers to clarify their queries. The user's input is obtained. If, in any case, the query is based on forecasts, the user will see a graphic representation of the future forecasts for the requested agricultural commodities.
- A crop forecasting system which was automated was developed by BRAC University [5].Before

farmers start the cultivating or growing procedure, the Android-based app recommends the most productive crops based on their land area. The profitability parameter completely covers the framework of the most efficient prompt crop [14]. Currently, the full system is being deployed for different Bangladeshi areas.

- In [1], a "no" response is returned by default if, in any case, the test case does not match with any keywords. According to [2], the concept is especially intended for regions like Kerala, where the average farm size is significantly smaller compared to any other Indian states.
- According to [5], the biggest drawback of the whole system created here was that it was solely built for the Android platform, meaning that users of Windows, iOS or any such system users might be unable to use it [15].

II. DATA COLLECTION AND DATASETS

1. Factors influencing the establishment of farming

a.) Effect of Rainfall– Rain is the primary source of agricultural energy in Maharashtra. As a result, precipitation has a big impact on the harvest.

b.) Effect of Temperature – A number of factors of crop growth are impacted by temperature. The ability of the plant to undertake photosynthesis and transfer water can change with low temperatures, whilst mineral enrichment, shoot growth, as well as pollen production are all negatively impacted by high temperatures. All of these elements influence crop yield either directly or indirectly.

2. Factors affecting the profitability of a crop

a.) Area under cultivation – The harvested output per unit of harvested area is used to determine crop yields. Therefore, whether a crop will turn out to be beneficial for the farmer depends on the harvested area.

b.) The Previous crop yield – To investigate seasonal variability and fluctuations in agricultural output for each season, historical data are needed.

3. Data Collection

a.) Crop statistics data - The website https://data.gov.in/catalog/district-wise-season-wise-crop-production-statistics was used to gather information for the previous year's crop statistics. It all revolves around the well- known platform of India, OGD (open government data) . For benefitting the general public, On the website, data sets, services, and apps are published by departments of the Indian government and their organisations. It contains information from 1997 to 2014. [5] [6] [10]. Screenshots from a database including data on district-level crop yield for several regions of Maharashtra are exhibited in Fig. 3.

b.) Rainfall Data: The website maharain.gov.in has provided data on rainfall in different parts of Maharashtra. The Department of Agriculture in Maharashtra is responsible for the web portal. In order to monitor daily rainfall at the divisional, district, and circle levels, it was first introduced in 1998. It contains information from 1997 to 2016.

4. Data Preprocessing

Before supplying raw data to the algorithm, a technique known as data preparation is utilized to transform it into a desirable and intelligible format. Data that is collected from the real world frequently have missing numbers, mistakes, and inconsistencies.

	A	B	C	D	E	F	G
1	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
2	Maharashtra	AHMEDNAGAR	1997	Autumn	Maize	1	1113
3	Maharashtra	AHMEDNAGAR	1997	Kharif	Arhar/Tur	17600	6300
4	Maharashtra	AHMEDNAGAR	1997	Kharif	Bajra	274100	152800
5	Maharashtra	AHMEDNAGAR	1997	Kharif	Gram	40800	18600
6	Maharashtra	AHMEDNAGAR	1997	Kharif	Jowar	900	1100
7	Maharashtra	AHMEDNAGAR	1997	Kharif	Maize	4400	4700
8	Maharashtra	AHMEDNAGAR	1997	Kharif	Moong(Gr	10200	900
9	Maharashtra	AHMEDNAGAR	1997	Kharif	Pulses tota	451	130
10	Maharashtra	AHMEDNAGAR	1997	Kharif	Ragi	2600	2100
11	Maharashtra	AHMEDNAGAR	1997	Kharif	Rice	5900	7200
12	Maharashtra	AHMEDNAGAR	1997	Kharif	Sugarcane	45900	38940
13	Maharashtra	AHMEDNAGAR	1997	Kharif	Total food	3384	1836
14	Maharashtra	AHMEDNAGAR	1997	Kharif	Urad	1600	800
15	Maharashtra	AHMEDNAGAR	1997	Rabi	Jowar	598400	217000
16	Maharashtra	AHMEDNAGAR	1997	Rabi	Maize	6200	9100
17	Maharashtra	AHMEDNAGAR	1997	Rabi	Other Rab	60	3
18	Maharashtra	AHMEDNAGAR	1997	Rabi	Wheat	79700	87100
19	Maharashtra	AHMEDNAGAR	1997	Summer	Maize	1100	1900
20	Maharashtra	AHMEDNAGAR	1997	Whole Yea	Cotton(lint	98	17900

Fig. 3. Data set of past-year crop statistics

The actions that are taken during data preprocessing are as follows:

a. Missing Values: The occurence of the categorical categories, averaging the data points, and other techniques can all be used to analyse missing values in data.
b. Data Normalisation: It is vital to normalise the data in order to prevent this from happening when the attributes in the data have different scales, as some features with a wider range may have a greater impact on the prediction. It is beneficial for distance-based prediction algorithms like K-Nearest Neighbors.

III. ALGORITHM AND IMPLEMENTATION

The following algorithms used in forecasting crops are :

1. KNN: K-nearest neighbor algorithm

a.) KNN algorithm - In comparison to other algorithms, K Nearest Neighbor is the most popular. The testing step makes use of all the training data that is stored. It classifies recent testing data using a measure of similarity with a distance function. It makes use of the K, or nearest neighbours, statistic. According to Fig. 4, the method locates the new point's closest neighbour and if somehow the class of the newly added point needs to be identified, gives its label to it and k's value is equivalent to 1. If value of k is greater than 1, them label is chosen by majority vote from its k neighbours, who then determine the k closest neighbours. To determine which point is the nearest to the other, the distance between them is computed using various distance formulas, including the euclidean distance and minkowski distance and many such other formulaes [19] [20] [21] [22].

b.) Modelling – K's value affects how accurate the model is. Because it catches the local points, the risk of overfitting grows as the value of k lowers. This will result in a high training accuracy but a low testing accuracy. If value of k increases and gets closer to the total no. of data points, the designed model will be underfitted since everything will be expected to belong to the one class that makes up the majority of the dataset [16] [17] [18].

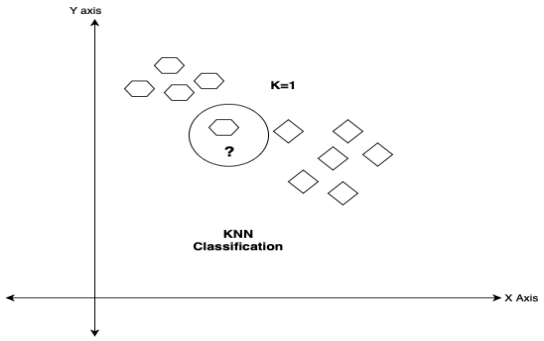


Fig. 4. KNN

2. Decision Tree

a.) A decision tree's operation: An approach for supervised machine learning called a decision tree is frequently employed for classification issues. Both categorical and continuous input variables can be used with it. This uses a significant splitter to divide the samples into homogeneous groupings. We adopt a categorical variable decision tree since our goal variable, the crop to be predicted, has a categorial value [14][15].

2) Splitting strategy: Entropy is the method employed in the splitting process. It refers to how disorganized a system is. The sample is said to be entirely homogenous if its entropy is zero, and it has an entropy of one if it is evenly divided. The formula below can be used to find out the entropy:-

Entropy = $-plog2p - qlog2q$

Here, the node's chances of success and failure are denoted by p and q, respectively. The decision tree is modeled in accordance with the split that has the lowest entropy. The performance of the model improves with decreasing entropy [15].

The dataset's total sample count is represented by the root node in Fig. 5, which is then further subdivided into homogenous sets. The algorithm separates the samples into further sub-nodes after

determining the most significant variable. Decision nodes are sub-nodes that divide into other sub-nodes. The nodes that cannot split are known as terminal nodes [14] [15] [23].
3)Modelling: Overfitting is often a problem for decision trees. Coaching accuracy would be 100% if the decision tree modelling settings are unrestricted. This might happen because it makes a leaf for each observation. However, the testing accuracy in such modelling is poor. It is necessary to take the following actions to prevent overfitting:

1) Limiting the size of trees
2) Pruning trees

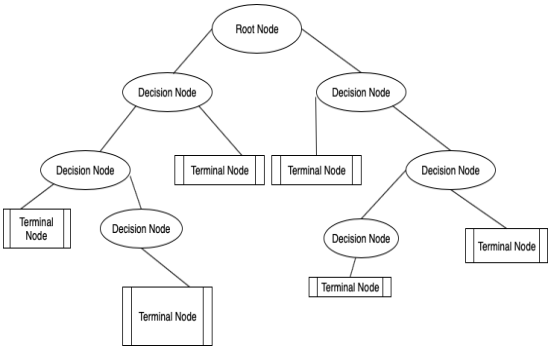


Fig. 5. Decision Tree terminologies

3. Random Forest Algorithm

a.) Working of Random Forest - Like Decision Tree, which grows many trees, Random Forest is also a supervised machine learning technique. Each tree inside this scenario will produce a single class and forest chooses the class with majority votes. [16].

It functions in the following manner –

- If there are "N" instances in the training set, then the sample 'N' instances is chosen randomly (with replacement). This will serve as the tree's training set [16].
- Assuming that there exists "M" input variables, "a" variables are then randomly chosen from M in such a way that "a" variable is smaller than M variable. The node is thought to be split using best split [16].
- Without any pruning, each tree is allowed to reach its full potential.
- Assemble or combine the forecasts of the n parent trees to forecast new test data [16]. The majority vote is used for categorization, whereas the average is used for regression, as demonstrated in fig.6.

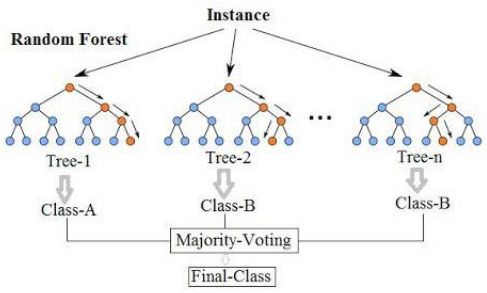


Fig. 6. Random Forest

IV. RESULT ANALYSIS

Both algorithms' outputs have been evaluated and set apart from one another. Table of results summarises the training and testing accuracies produced by the various algorithms. The

following topics further highlighted by an evaluation of the algorithms and respective results are :

1) KNN –
   - Large training data yields effective results.
   - Label encoding required(in case of categorical values).
   - Estimation of K's value.
   - Expensive computation owing to distance calculations for each test result.

Used:

```
#import RandomForestClassifier

from sklearn.ensemble
import RandomForestClassifier

#Fit the data
classifier_rf.fit(X_train, y_train)
```

```
#import KNNClassifier

from sklearn.neigbors
import KNeighborsClassifier
from sklearn import metrics

#Fit data
Classifier =
KNeighborsClassifier(n_neighbors=
5, metric='minkowski', p=2 )
classifier.fit(x_train, y_train)
```

2) Decision Tree –
   - Depiction by graph is highly intuitive.
   - It requires minimal data cleaning and can handle categories and numerical variables. Missing values can also be handled by it.
   - If no parameters are constrained, it may result in overfitting.

Used:

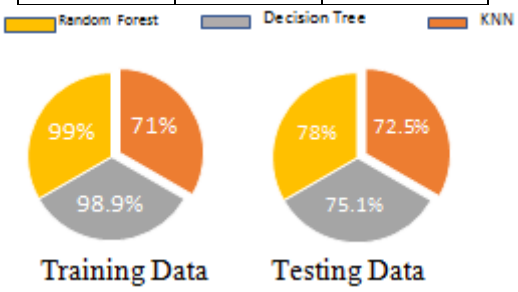
```
#import DecisionTreeClassifier

from sklearn.tree
import DecisionTreeClassifier

#Fit the data
clf = tree.DecisionTreeRegressor()
clf = clf.fit(X, y)
```

3)
   - Able to handle massive, higher-dimensional data sets.
   - Fills up the gaps in the data while maintaining accuracy.
   - For statistical modellers, comparable to a black box method (lesser control over other parameters)

Used:

TABLE OF RESULTS
ANALYSIS OF THE ALGORITHMS

Algorithm	Accuracy - Training	Accuracy - Testing
1.KNN	71%	72.5%
2.Decision Tree	98.9%	75.1%
3.Random Forest	99%	78%



V. CONCLUSION

The main goal of the entire project was to give farmers a virtual farming assistant that can engage with them and speak with them. Even in the age of digital disruptions, farmers continue to face many difficulties. The software was created specifically to assist farmers with all the nuances of farming and make it very farmer-oriented. Farm-assist not just allows farmers get the best recommendations in crops, but also helps them take good care for their crops and store them to lengthen their shelf life, and locate the best prices for trading their commodities in neighbouring marketplaces, all of which have a beneficial economic impact.

VI. REFERENCES

[1] S.Pudumalar, E.Ramanujam, R.Harine Rajashreen, C.Kavyan, T.Kiruthikan, J.Nishan Thiagarajar College of Engineering," Crop Recommendation System for Precision Agriculture", IEEE Eighth International Conference on Advanced Computing, March 2016

[2] Satish Babu, "A Software Model for Precision Agriculture for Small and Marginal Farmers", The International Centre for Free and Open Source Software (ICFOSS) Trivandrum, India, May 2013

[3] Anshal Savla, Parul Dhawan, Himtanaya Bhadada, Nivedita Israni, Alisha Mandholia , Sanya Bhardwaj," Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture", Innovations in Information, Embedded and Communication systems (ICIIECS), May 2015

[4] Mrs. L. Kannagi, Ramya .C, Shreya.R , Sowmiya R Valliammai Engineering College, " Virtual Conversational Assistant The FARMBOT", IJETSR, ISSN 2394, Volume 5, Issue 3, March 2018.

[5] Talha Siddique, Dipro Barua, Zannatul Ferdous, Amitabha Chakrabarty, BRAC University," Automated Farming Prediction", Intelligent Systems Conference 2017 ,7-8 September 2017 London, UK

[6] Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh," Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique", International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM),2015

[7] Monali Paul, Santosh K. Vishwakarma, Ashok Verma," Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach", International Conference on Computational Intelligence and Communication Networks, India, May 2015

[8] V. Kumar, V. Dave, R. Bhadauriya, S. Chaudhary," Krishimantra:" agricultural recommendation system", Proc. of the 3rd ACM Symposium on computing for development, no. 45, 2013.
[9] J. Ranjan," Data Mining Techniques for better decisions in Human Resource Management Systems", International Journal of Business Information Systems IJBIS, vol. 3, pp. 464-481, 2008.
[10] Paul Monali, Santhosh K. Vishwakarma, Ashok Verma," Prediction of crop yield using Data Mining approach", Computational Intelligence and Communication Networks (CICN) International Conference, 12-14 Dec. 2015
[11] Narayan, V., & Daniel, A. K. (2022). FBCHS: Fuzzy Based Cluster Head Selection Protocol to Enhance Network Lifetime of WSN. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, *11*(3), 285-307.
[12] Narayan, V., & Daniel, A. K. (2019). Novel protocol for detection and optimization of overlapping coverage in wireless sensor networks. *Int. J. Eng. Adv. Technol*, *8*.
[13] Awasthi, S., Srivastava, A. P., Srivastava, S., & Narayan, V. (2019, April). A Comparative Study of Various CAPTCHA Methods for Securing Web Pages. In *2019 International Conference on Automation, Computational and Technology Management (ICACTM)* (pp. 217-223). IEEE.
[14] Narayan, V., & Daniel, A. K. (2022). Energy Efficient Protocol for Lifetime Prediction of Wireless Sensor Network using Multivariate Polynomial Regression Model. *Journal of Scientific & Industrial Research*, *81*(12), 1297-1309.
[15] Choudhary, S., Narayan, V., Faiz, M., & Pramanik, S. (2022). Fuzzy approach-based stable energy-efficient AODV routing protocol in mobile ad hoc networks. In *Software Defined Networking for Ad Hoc Networks* (pp. 125-139). Cham: Springer International Publishing.
[16] Srivastava, S., & Sharma, S. (2019, January). Analysis of cyber related issues by implementing data mining Algorithm. In *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (pp. 606-610). IEEE.
[17] Srivastava, S., Yadav, R. K., Narayan, V., & Mall, P. K. (2022). An Ensemble Learning Approach For Chronic Kidney Disease Classification. *Journal of Pharmaceutical Negative Results*, 2401-2409.
[18] Srivastava, S., & Singh, P. K. (2022). Proof of Optimality based on Greedy Algorithm for Offline Cache Replacement Algorithm. *International Journal of Next-Generation Computing*, *13*(3).
[19] Salagrama, S., Kumar, H. H., Nikitha, R., Prasanna, G., Sharma, K., & Awasthi, S. (2022, May). Real time social distance detection using Deep Learning. In *2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES)* (pp. 541-544). IEEE.
[20] Mahadani, A. K., Awasthi, S., Sanyal, G., Bhattacharjee, P., & Pippal, S. (2022). Indel-K2P: a modified Kimura 2 Parameters (K2P) model to incorporate insertion and deletion (Indel) information in phylogenetic analysis. *Cyber-Physical Systems*, *8*(1), 32-44.
[21] Tyagi, N., Rana, A., Awasthi, S., & Tyagi, L. K. (2022). Data Science: Concern for Credit Card Scam with Artificial Intelligence. In *Cyber Security in Intelligent Computing and Communications* (pp. 115-128). Singapore: Springer Singapore.
[22] Awasthi, S., Kumar, N., & Srivastava, P. K. (2021). An epidemic model to analyze the dynamics of malware propagation in rechargeable wireless sensor network. *Journal of Discrete Mathematical Sciences and Cryptography*, *24*(5), 1529-1543.
[23] Singh, M. K., Rishi, O. P., Awasthi, S., Srivastava, A. P., & Wadhwa, S. (2020, January). Classification and Comparison of Web Recommendation Systems used in Online Business. In *2020 International Conference on Computation, Automation and Knowledge Management (ICCAKM)* (pp. 471-480). IEEE.