



Farmright – A Crop Recommendation System

Dviti Arora^(✉), Sakshi, Sanjana Drall, Sukriti Singh, and Monika Choudhary

Computer Science and Engineering, Indira Gandhi Delhi Technical University for Women,
Delhi, India

{dvitiarora125btcse18, sakshi097btcse18, sanjana107btcse18,
sukritisingh111btcse18, monikachoudhary}@igdtuw.ac.in

Abstract. Agriculture is extremely vital to our economy and boosting the development of this sector always adds up to the economic & political value of our country. Health of all the crops grown is affected by various aspects including technological, biological, and environmental factors. The environmental facet particularly has been drastically changing, posing challenges in front of the peasants. They face a significant difficulty in determining the optimal crop for their farming region to maximize productivity and profit. For Indian farmers, there is no existing reliable recommendation mechanism. Giving an address to this issue, the study proposes a crop recommendation system based on a multi-label classification model which considers the location of peasants, composition of soil, and weather characteristics, and provides a ranked list of suggested crop seed to be produced for greater yield. Researchers compare many algorithms based upon the performance criteria and capabilities to develop the best recommendation model for crops. With a precision of 82.74%, a recall of 80.92%, and an F1 score of 78.67%, the most optimal model was revealed to be an RF Technique. The trained model proved advantageous in catering the farmers with a ranked list of crops deployed along with an interface for better user experience.

Keywords: Crop recommendation · Agriculture · Multi-label classification · Random Forest Algorithm · Soil characteristics · Climate properties · Chatbot

1 Introduction

Agronomy and cultivation are a crucial part of our economic sector acting as the lifeblood of rural households with 70% [20] of them depending on agriculture for their livelihoods. This sector accounts for about 13.7% [11] of India's total GDP. It is the largest employment sector that employs more than 57% [11] of the population making up about two-thirds of the working population in India. It also impacts international trade, food grain supply, and input to other industrial sectors. This increasing demand can only be met by increasing agricultural production. So, there is raising pressure on the country's smallholders to meet the demand for food grain requirement. With that in mind, we focus on dramatic changing factors such as weather and soil that affect plant health. Farm owners today confront considerable challenges in picking the best crop for their sowing acreage in order to increase output and profit. They need to choose the best

harvest for their land. They have always tried to pick desirable characteristics in seed and harvest for many years, upgrading the plant quality and quantity for agricultural reasons. They started looking for crop types with shorter growing seasons, larger seeds, and fruits. Agricultural technology has given rise to a diverse range of food, fruit, feed, and fiber production options. In all manner of ways, technology makes our life easier and more fun by lowering the amount of efforts and time we spend on necessary tasks such as food production. Farmers sometimes fail to select the appropriate crops and sowing locations depending on environmental considerations. This became a research area to heighten the knowledge and improve agricultural technology. Studies tried to improve the crop selection methods & emulsifying it with technology. The selection method requires the latest data & knowledge about agricultural tendencies which lacked in recent studies. Since there are myriad factors affecting the process of growing a crop, the complexity of agricultural technology proliferates. For Indian farmers, there is no official trustworthy agricultural advisory system as of now. If you choose the wrong crop or seeding area, your investment, resources, and efforts will be wasted. An equally notable problem these days is the decreasing proportion of farmers who cultivate large areas. Farmers are afraid of mass planting due to unpredictable yields, erratic weather, and lack of assets. Keeping this in view, researchers have been trying to devise ways that can help mitigate the problems and boost the sector. Various technologies are used in the process with Machine Learning [7] being the most used one. ML models are trained using the agricultural data like temperature, rainfall, soil type and properties, traditionally grown crops in the concerned area, production per unit area of previous decades so that they can recommend the most suitable crop.

This paper aims at proposing a crop recommendation system called FarmRight, to tackle such problems. The system takes in parameters that affect the growth of crops as input and recommends the most favourable crop as output using Machine Learning algorithms [3]. The recommendation system employs a supervised learning task viz. Multi-label classification in which an occurrence in the training data is linked to one or more mutually non-exclusive class labels. To determine the best accurate crop recommendation model, researchers compare multiple algorithms based on their performance criteria. A Random Forest Algorithm was shown to be the best accurate model, with a precision of 82.74 percent, recall of 80.92 percent, and an F1 score of 78.67 per cent. The model trained up to this point proved advantageous in catering the farmers with a ranked list of recommended crops.

By the same token, this paper is organised along these lines, in Sect. 2, a brief review towards some related work on crop recommendation systems is presented. Section 3 contains the details about the solution that the paper aims at proposing, followed by Sect. 4 which explains the steps followed for making the model. Implementation results are presented and discussed in Sect. 5. Section 6 concludes and tells about the future scope of the paper.

2 Background and Literature Review

GiVen the importance of agriculture in our economy, researchers have been trying to come up with methods that would help this sector boost. Various techniques have

been used starting from use of sensors to ensembling approach of Machine Learning. Vijayabaskar et al. [19] develops a project to determine soil fertility and proposes that crops should be planted based on sensor readings. Users can share information and receive assistance from specialists in a “farmer chat” created by the writers. It allows farmers to assess their land’s fertility and plant the most productive and profitable crop. It also offers information on the type of fertilizer to use on the soil as well as the location of a fertilizer store nearby. Patel et al. [13] collected data from Indian Government Soil testing card website and experimented on the data using R programming to build a Land Recommendation System (LSR) which recommends crops predicted on different parameters. Indira et al. [4] proposed a recommendation system based on machine learning to increase agricultural yield. Multiple algorithms used include MobileNet, which identifies disease using image leaf, XGBoost which suggests crop on the basis of local rainfall and soil properties and Random Forest to predict fertilizer values. Prakash et al. [18] collected soil moisture, pH, temperature and humidity values from field with the help of sensors and sent to cloud using Node MCU. Data in cloud was used to analyze different ML algorithms and compared their prediction accuracy. S Pudumalar et al. [14] developed a crop recommendation system using the technique of data mining that will assist cultivators in choosing the appropriate crop based on soil requirements to boost production and profit. Precision farming is an improved cultivation approach that employs explored datasets on soil properties, soil type, and yield data to advise cultivators on the suitable crop for their sowing area. This solution is a system of recommendation that uses algorithms such as CHAID, KNN, Naive Bayes etc. as learners. Banerjee et al. [1] build an proficient and potent crop recommendation system taking in account various parameters such as soil type and nature, terrain and rainfall pattern using fuzzy logic for West Bengal. Liu et al. [10] the author suggests a clustering centre optimised technique using the Synthetic Minority Over-sampling Technique (SMOTE) to experiment on imbalanced soil data. The procedure begins by examining the original sample points and selecting density based grouping centres. The clustering centre is then used to generate minority samples in order to balance out the data distribution. Finally, for accurate prediction, the ensemble approach is employed to train the prediction model. Kumar et al. [8] suggested the Crop Selection Method (CSM) as a method for solving crop selection problems which also uses machine learning techniques. The proposed strategy has the potential to increase crop net production rates. Sujjaviriyasup et al. [16] proposed SVM model for forecasting exports of agricultural products and ARIMA was utilised as a benchmark to assess other developed models, and Thailand’s Pacific white shrimp export data and data on chicken produced in Thailand were used. Ramya et al. [15] aims at providing a Climate Smart and agriculture decision-making platform for changing climate that affects agriculture. Using data and analytics, they were able to predict and mitigate the impact of catastrophic weather events on global finance and the Economic Dimensions of Climate Change all over the world. Kulkarni et al. [9] projected that a crop recommendation system based on the ensembling approach of ML would be built to boost agricultural productivity. It aims at designing a system for recommending the precise selection of seed based on multiple soils and climatic factors is the goal of this research work. The purpose of the research is to increase crop productivity by using the ensembling technique to provide high-accuracy and efficient predictions. The

research work presented by Teja et al. [17] includes a website integrated with ML algorithms such as K-Nearest Neighbour (KNN), Naïve Bayes, SVM algorithm, etc., and historical weather data to find the most profitable crop under current weather conditions. As Machine Learning algorithms advanced, Jeong et al. [5] investigated RF regression effectiveness and predicted complicated crop yield responses. They discovered that RF performs well in predicting yields for all plants and areas studied. The findings of this study suggest that RF algorithms have a lot of potential as a statistical modeling tool for yield prediction. In every performance measure evaluated, RF results outperformed the MLR benchmark. The most significant observation was the mean square error of all test cases in the RF model ranged from 6 percent to 14 percent of the mean observed yield, while these values ranged from 14 percent to 49 percent of the compared model. The study of Banjara et al. [2] shows how diversifying the rice-wheat crop system can improve the rice produce and increase profits in Indo-Gangetic Plain region. Rice-berseem-cowpea fodder and rice-potato-green gram were found the most appropriate sequence of cropping. This was found with the help of nutrient content analysis and calculating nutrient uptake by grain and straw. Krishna et al. [6] studies and examines crop-climate relationships in India using historical production data for major crops (rice, groundnut, wheat, sugarcane, and sorghum), as well as aggregate pulses, food grain, oil seed and cereal production.

Above said approaches worked well but some limitations were tagged along such as the set of crops they could predict was limited to few crops only as in [9] though there are more than hundreds of crops that are grown in India. Moreover, dataset used focused on data from few locations as in [14] and [19]. Algorithms used and the accuracy of the model were not mentioned clearly as in [19].

3 Proposed Solution

Agriculture-DEPENDENT countries, such as India, rely on agriculture for economic growth. As the country's population grows, so does its reliance on agriculture, which impairs subsequent economic growth. Therefore, it is necessary to increase the yield of crops. Some biological procedures along with a few chemical techniques can find solutions to this problem. Additionally, a crop recommendation model is needed to improve seasonal crop yield by recommending a list of acceptable crops based on location, soil, and climatic parameters. To improve crop productivity and facilitate correct decision-making for the selection of crops, this project proposes a recommendation system that recommends a list of apt crops based on the location, climate, and field properties using Multi-Label Classification techniques. The selection of crops is dependent on various parameters. The datasets used have been collected based on these affecting factors and include the soil properties (Macronutrients and soil pH levels) for all districts in India, the agro-climatic factors such as precipitation and temperature, and the district wise crop distribution for all states of India. These datasets obtained from different sources have been merged to create one master dataset. The master dataset contains instances associated with about 124 different crops ranging from food and non-food crops to crops grown in different seasons in different parts of India. This dataset obtained is then used to train different machine learning classification models using two multi-label classification techniques, Binary Relevance and Classifier Chain. To find the most appropriate

model, different performance metrics corresponding to each model such as recall, precision, f1 score, and hamming loss are compared. For the crop recommendation system, the model with the lowest hamming loss and significantly high precision, recall, and f1 score is preferred. After finding the best possible model, it can be used by the users to give a recommended list of crops if they enter their location, soil characteristics of their land, and the climatic variable such as rainfall and temperature. This system can help in giving the farmers/ users a more personalized experience in deciding crops according to their location and land parameters. It also provides a user-friendly way to assist in the right crop selection decisions. Another benefit of the system proposed is that it doesn't restrict its results to 1 or 2 crops, but with the help of multi-classification techniques gives a wide range of crops to choose from. Our model can potentially predict around 124 crops (Fig. 1).

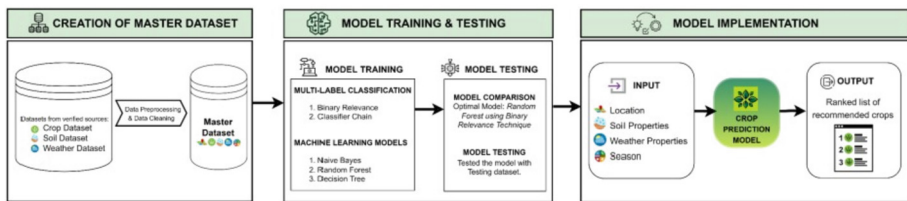


Fig. 1. Workflow diagram

4 Methodology

4.1 Data Collection

THERE are four categories into which major crops in India can be divided into viz. Food Grains such as Wheat, Maize, Rice, Pulses, Millets, etc., Cash Crops including Jute, Cotton, Sugarcane, Oilseeds and Tobacco, Plantation Crops such as Coconut, Rubber, Coffee and Tea and Horticulture crops including Fruits and Vegetables. The yield potential of the crops, in particular, depends on the soil characteristics of the land and the climate. Adequate nutrition of plants is critical for the optimal production of crops. Primary macronutrients are very vital for the yield quality enhancement of crops. Nitrogen (N), phosphorous (P), and potassium (K) are the three primary elements that must be consumed in large quantities. Soil pH is an important variable in soils because it regulates numerous chemical and biological activities. It refers to the soil's acidity or basicity/alkalinity. In agriculture, soil pH is important because it regulates plant nutrition availability by controlling the chemical forms of various nutrients and their chemical interactions[12]. As a result, the soil pH has a significant impact on soil and crop yield. The monsoon, which originates from the Indian and Arabian Seas, is vital to India's agriculture. As the climate changes, farmers are unprepared for changes in the rainfall cycle, magnitude, and timing. Soil moisture evaporates quickly in arid climates, reducing water available for agricultural production. Climate change has an impact on groundwater levels in different parts of the world.

Therefore, the following datasets were collected from different sources:

Crop Dataset [21]. This crop dataset contains the distribution of crops in different regions and the production statistics. The data can be utilized to study and analyze crop production, contribution to regional production, agricultural climate zone performance, crop growth pattern and high yield production for crops, and diversification. The dataset has approximately 246091 rows and 7 columns. Source of this dataset is Open Government Data (OGD) Platform India [22].

Soil Dataset [21]. This dataset sourced by the DoAFW [23] contains the percentage of Macro-nutrients in a particular district based on the year cycle. Macro Nutrients comprises Organic carbon(OC), Potassium(K), Nitrogen(N), and Phosphorus(P). It also includes the percentage of Acidity, Neutrality, and Alkalinity of the soil in a particular district year-wise. The pH is tested and categorized based on the acidic, basic, and neutral elements found in the soil, such as Acid Sulphate, Calcium, Magnesium, Sodium, etc. The idea is to take the average of acidic, basic, and neutral percentages for our dataset. Its approximate size is 748 rows and 8 columns.

Climate Dataset. The climate dataset contains monthly readings of a particular Geographic coordinate (District-wise) taken in a year. The Agro-Climatic parameters include: Temperature at 2 m (°C), Precipitation Corrected (mm). The dataset had around 14212 instances. This dataset was collected from API from NASA Prediction Of Worldwide Energy Resources (POWER) [24] (Fig. 2).

CROP DATA SET				CLIMATE		SOIL PROPERTIES							LABEL
State_Name	District	Season	Prod/Area	Temp	Rainfall	N	P	K	OC	Acidic	Neutral	Basic	Crop
ANDAMAN An	NICOBAR	Kharif	0.33167104	27.5725	9.145	99.11	78.16	89.89	0.6596	0	8.53	0	Arecanut
ANDAMAN An	NICOBAR	Kharif	0.33167104	27.5725	9.145	99.11	78.16	89.89	0.6596	35.27	8.53	56.2	Other Kharif
ANDAMAN An	NICOBAR	Kharif	0.33167104	27.5725	9.145	99.11	78.16	89.89	0.6596	35.27	8.53	56.2	Rice
ANDAMAN An	NICOBAR	Whole Year	0.22050956	27.5725	6.08	99.11	78.16	89.89	0.6596	35.27	8.53	56.2	Banana
ANDAMAN An	NICOBAR	Whole Year	0.22052956	27.57	6.08	99.11	78.16	89.89	0.6596	35.27	8.53	56.2	Cashewnut
ANDAMAN An	NICOBAR	Whole Year	0.22052956	27.57	6.08	99.11	78.16	89.89	0.6596	35.27	8.53	56.2	Coconut
ANDAMAN An	NICOBAR	Whole Year	0.22052956	27.57	6.08	99.11	78.16	89.89	0.6596	35.27	8.53	56.2	Dry ginger
ANDAMAN An	NICOBAR	Whole Year	0.22052956	27.57	6.08	99.11	78.16	89.89	0.6596	35.27	8.53	56.2	Sugarcane
ANDAMAN An	NICOBAR	Whole Year	0.22052956	27.57	6.08	99.11	78.16	89.89	0.6596	35.27	8.53	56.2	Sweet potato
ANDAMAN An	NICOBAR	Whole Year	0.22052956	27.57	6.08	99.11	78.16	89.89	0.6596	35.27	8.53	56.2	Tapioca

Fig. 2. Master dataset made by combining crop, soil and climate datasets

4.2 Data Preprocessing

THE second stage of the proposed solution was to preprocess the data and the steps followed were:

- Removing rows with missing values in crop dataset.
- Replacing missing values in columns with mean of the adjacent states in soil dataset as per the Map of India, 2021.
- Adding columns corresponding to each season for the climate dataset.
- Making state and district names consistent in all 3 datasets for merging.
- Label Encoding the categorical columns such as state, district, season, and One Hot Encoding the Crop Label.

- Merging the 3 datasets on the columns State, District, and Year and dropping duplicate and unrequired columns to get the Master Dataset.

This resulted in the creation of a master dataset which had attributes: State, District, Season, Temperature, Rainfall, percentage of macronutrients (N, P, OC, K), and pH values categorized into Acidic, Basic, and Neutral percentages present and crop name as the Label. Now, the master dataset has 242361 rows with 14 columns.

4.3 Multi-label Classification

When given an input, classification refers to a predictive modelling problem in which a class label is obtained as an output. This differs from regression situations in which a numeric value is predicted. In most classification problems, a single label is predicted. It could also refer to estimating the likelihood of two or more different class designations. In such certain cases, the classes are mutually exclusive, which means the classification task believes the input belongs to only one class. However, there are times when more than one class label needs to be predicted. In this sense, class labels are mutually nonexclusive. These tasks are known as multiple label classification. Multi-label categorization produces zero or more labels as an output for every sample of input. It is assumed that the output labels can be defined as a function of the inputs. In our system, since we aim to recommend a list of crops, i.e. we need multiple crop predictions, multi-label classification is an apt technique. For single-label classification tasks, most classical learning methods are designed. As a result, numerous approaches in the literature reduce the multi-label problem to a series of single-label problems that can be addressed using current single-label algorithms. Most traditional learning methods are developed for single-label classification tasks. As a result, numerous approaches in the literature reduce the multi-label problem to a series of single-label problems that can be addressed using current single-label algorithms. Here, we have considered two techniques of multi-label classification:

Binary Relevance. Here, an ensemble of single-label binary classifiers is used and trained for every single class. Every single classifier determines whether or not an instance belongs to a specific class. To produce the multi-label result, all predicted classes are considered simultaneously. This is a common strategy since it is straightforward to implement, but it ignores any potential relationships between class labels.

Classifier Chains. In this case, a chain of n binary classifiers is formed, where n is the number of possible class labels and each classifier uses the predictions of all its predecessors. This allows us to account for label correlations in this way.

4.4 Models Used

Since we had got labeled data, some of the supervised classification algorithms considered were:

Naive Bayes. The Naive Bayes classifier is a straightforward heuristic classifier that employs Bayes' theorem (from Bayesian statistics) and the strong naive independent hypothesis. This technique uses classifier models to assign class labels to examples that are represented as vectors of feature values, with the class labels coming from a finite set. It is a collection of algorithms rather than a single algorithm based on a common premise. All naive Bayes classifiers assume that the value of one attribute is independent of the value of any another attribute given the class variable. Because Naive Bayes is useful for multidimensional information, it is taken into account in this scenario. For our system, we have considered two Naive Bayes classifiers, namely Multinomial Naive Bayes and Gaussian Naive Bayes.

Decision Tree. Decision trees use a tree structure to gather information. For better clarity and understanding, this tree structure can alternatively be represented as a series of discrete rules. The flexibility to use different subsets of attributes and decision rules at different phases of categorization gives decision tree classifiers an advantage. A root node, internal and leaf nodes, and branches make up a standard decision tree. The leaf node represents the class that will be allocated to a sample. Each internal node in a tree represents an attribute, while branches represent feature combinations that lead to classifications. The decision tree divides the population of data into smaller segments, allowing for more accurate predictions.

Random Forest. RFA is a tree ensemble method-based machine learning bagging approach. It creates a series of trees based on randomly subsampled attributes. The result of RFA is the average value of individual tree projections. Because RFA uses random sub-sampled features, it can be used in high-dimensional input predictors.

4.5 Model Training and Performance Metrics

THE dataset has been split into training and testing samples in the ratio 70:30. The partitioning of the dataset and the availability of testing data aid in the evaluation of the model's performance. The training data is then given as input to each of the classifiers independently. After successfully training and testing, the following evaluation metrics were considered to evaluate the performance of the models and for the comparison of their performance:

Hamming Loss. Hamming-Loss is an example-based metric that measures the percentage of labels that are wrongly predicted or the proportion of incorrect labels to the total number of labels.

Precision. It is the propensity of a classification model to identify only the pertinent data points. It's strategically calculated by dividing the count of genuine positives by the total number of true positives plus false positives.

Recall. A classification model's capacity to find all relevant examples bounded in a data collection. The count of true positives divided by the count of true positives added to the count of false negatives is called recall.

F1 Score. The F1 Score is a means to integrate precision and recall into a single metric that accounts for both. It's the harmonic mean of the precision and recall of a classifier's predictions, therefore it's an overall measure of the quality of the prediction.

5 Results and Discussions

To find the performance of our models, we tested the models using the testing data. The performance metrics were observed for each of the models. It was observed that both the Naive Bayes algorithms, Multinomial Naive Bayes and Gaussian Naive Bayes had extremely low F1 scores, indicating that the model wasn't performing very well. Whereas, the Random Forest Algorithm using the Binary Relevance technique returned the least hamming loss and high values of recall, precision, and F1 score. RFA proved to be efficient for our use case due to the very large dataset. Since we have a large dataset, and a variety of features, a lot of rules can be induced from it. Random forest has the ability to discover all the rules in the dataset as it creates multiple trees while training, hence it is able to cover all existing rules/patterns in the dataset, thus giving better accuracy. We also observe that binary relevance technique gives the most optimal results. Being the most simple multi-label classification technique, it is easily parallelized for our large dataset. It only needs one-step learning and is able to optimize several loss functions. Hence, it can be concluded that the RFA with Binary Relevance is the most suited algorithm for our proposed solution (Tables 1 and 2).

Table 1. Performance evaluation for binary relevance

Model	Hamming Loss	Precision	Recall	F1-Score
Multinomial NB	30.02	11.39	61.64	17.26
Gaussian NB	07.72	37.74	55.96	39.34
Decision Tree	02.76	77.22	79.25	75.15
Random Forest	02.23	82.72	80.92	78.76

Table 2. Performance evaluation for classifier chain

Model	Hamming Loss	Precision	Recall	F1-Score
Multinomial NB	39.95	08.26	42.75	12.49
Gaussian NB	18.97	18.15	63.13	26.13
Decision Tree	02.80	77.98	78.33	74.49
Random Forest	02.49	82.00	79.89	77.35

FarmRight, the crop recommendation model, is further integrated with a website to provide an interface to the users for greater accessibility and usage. Farmers can visit the “Farm Right” website and get to know which is the most suitable crop for his land by filling details about the farmer’s land and other properties. It also caters with a Load Properties button that auto-fills the common soil and weather properties’ values from the database, to abstract the users from the technical details such as pH and NPK values. It uses the built crop-recommendation model FarmRight to recommend a list of crops as an output. The website provides a brief detail of the problems faced by farmers of our country (Figs. 3 and 4).



Fig. 3. FarmRight Website

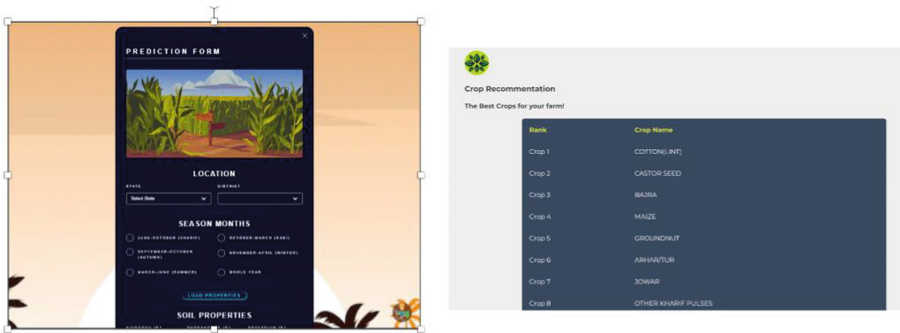


Fig. 4. User Input and Output

Along with this it has a chatbot “KrishiBot” which enables users to explicitly ask their queries regarding chat recommendation system and the workflow and get the best answers. The chatbot is implemented using NLP using a set of intents and responses that helps to answer queries asked by the user (Fig. 5).

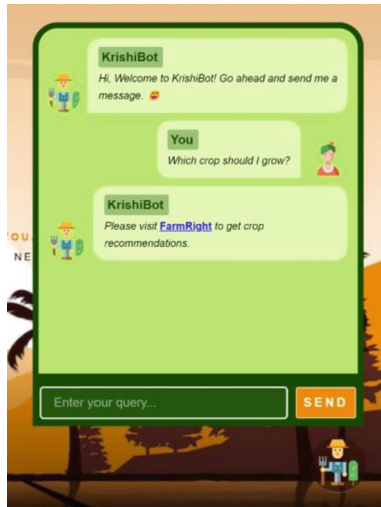


Fig. 5. KrishiBot

6 Conclusion and Future Scope

India is a country where agriculture plays a vital part. The country prospers in the prosperity of the peasants. As a result, our study assists farmers in sowing the appropriate seeds based on soil requirements and climatic factors in order to boost productivity and benefit from strategies like these. This allows farmers to plant the right crops to increase yields and increase the overall productivity of the land. The project aims at implementing multi-label classification on the collected dataset using different algorithms to find the best fit algorithm for the recommendation system. To determine the best accurate model, paper compares different algorithms based on their performance criteria. A Random Forest Algorithm was discovered to be the most accurate model, with a precision of 82.74 percent, a recall of 80.92 percent, and an F1 score of 78.67 percent. The trained model proved beneficial in assisting the farmers with a ranked list of crops deployed along with an interface for better user experience.

In future, recently updated datasets with more attributes like fertilizers added and humidity could be used for prediction of crops. We can further look for other model building techniques and model tuning like hyperparameter tuning can be performed as well. Also, the website could be made multilingual, connecting to experts and helpline for scalable use. The chatbot capabilities can be enhanced to handle extensive and complex queries for better user experience. Along with this, fertilizer recommendation, crop disease detection and profit made, depending upon region and crop chosen by the user can be added as another feature in the website.

References

1. Banerjee, G., Sarkar, U., Ghosh, I.: A fuzzy logic-based crop recommendation system. In: Proceedings of International Conference on Frontiers in Computing and Systems, pp. 57–69. Springer (2021)
2. Banjara, T.R., Bohra, J.S., Kumar, S., Ram, A., Pal, V.: Diversification of rice–wheat cropping system improves growth, productivity and energetics of rice in the Indo-Gangetic plains of India. *Agric. Res.* **11**(1), 48–57 (2021)
3. Elavarasan, D., Vincent, P.M.D.R.: A reinforced random forest model for enhanced crop yield prediction by integrating agrarian parameters. *J. Ambient. Intell. Humaniz. Comput.* **12**(11), 10009–10022 (2021). <https://doi.org/10.1007/s12652-020-02752-y>
4. Indira, D.N.V.S.L.S., Sobhana, M., Swaroop, A.H.L., Phani Kumar, V.: KRISHI RAKSHAN - A Machine Learning based New Recommendation System to the Farmer. In: 6th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1798–1804. IEEE Xplore (2022)
5. Jeong, J.: 2016 Random forests for global and regional crop yield predictions *PLoS ONE* **11**(6), e0156571 (2016)
6. Krishna Kumar, K., Rupa Kumar, K., Ashrit, R., Deshpande, N., Hansen, J.: Climate impacts on Indian agriculture. *Int. J. Climatol.: J. R. Meteorol. Soc.* **24**(11), 1375–1393 (2004)
7. Kumar, R., Singhal, V.: IoT enabled crop prediction and irrigation automation system using machine learning. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)* **15**(1), 88–97 (2022)
8. Kumar, R., Singh, M., Kumar, P., Singh, J.: Crop selection method to maximize crop yield rate using machine learning technique. In: 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), pp. 138–145. IEEE (2015)
9. Kulkarni, N., Srinivasan, G., Sagar, B., Cauvery, N.: Improving crop productivity through a crop recommendation system using ensembling technique. In: 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), pp. 114–119. IEEE (2018)
10. Liu, A., Lu, T., Wang, B., Chen, C.: Crop recommendation via clustering center optimized algorithm for imbalanced soil data. In: 2020 5th International Conference on Control, Robotics and Cybernetics (CRC), pp. 31–35. IEEE (2020)
11. Malik, A., Kumar, R.: An overview on agriculture in India. *Int. J. Mod. Agric.* **10**(2), 2087–2095 (2021)
12. Oduola Oshunsanya, S.: Introductory Chapter: Relevance of Soil pH to Agriculture. *Soil pH for Nutrient Availability and Crop Performance*, IntechOpen, London (2019). <https://doi.org/10.5772/intechopen.82551>
13. Patel, K., Patel, H.: A state-of-the-art survey on recommendation system and prospective extensions. *Comput. Electron. Agric.* **178** 105779 (2020)
14. Pudumalar, S., Ramanujam, E., Rajashree, R., Kavya, C., Kiruthika, T., Nisha, J.: Crop recommendation system for precision agriculture. In: 2016 Eighth International Conference on Advanced Computing (ICoAC), pp. 32–36. IEEE (2017)
15. Ramya, M., Balaji, C., Girish, L.: Environment change prediction to adapt climate-smart agriculture using big data analytics. *Int. J. Adv. Res. Comput. Eng. & Technol. (IJARCET)* **4**(5) (2015)
16. Sujjaviriyasup, T., Pitiruek, K.: Agricultural product forecasting using machine learning approach. *Int. J. Math. Anal.* **7**(38), 1869–1875 (2013)
17. Teja, M.S., Preetham, T.S., Sujihelen, L., Christy, Jancy, S., Selvan, M.P.: Crop recommendation and yield production using SVM algorithm. In: 6th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1768–1771 (2022)

18. Varun Prakash, R., Mohamed Abrith, M., Pandiyarajan, S.: Machine learning based crop suggestion system. In: 6th International Conference on Intelligent Computing and Control Systems (ICICCS), pp. 1355–1359. IEEE Xplore (2022)
19. Vijayabaskar, P., Sreemathi, R., Keertanaa, E.: Crop prediction using predictive analytics. In: 2017 International Conference on Computation of Power, Energy Information and Communication (ICCPEIC), pp. 370–373. IEEE (2017)
20. India at a Glance, FAO in India. <https://www.fao.org/india/fao-in-india/india-at-a-glance/en/>. Accessed 28 Jan 2022
21. FarmRight, Github. <https://github.com/Know-and-Grow/FarmRight-A-Crop-Recommendation-System>. Accessed 31 Aug 2022
22. Open Government Data (OGD) Platform India. <https://data.gov.in/>. Accessed 10 Feb 2022
23. Department of Agricultural Cooperation & Farmers Welfare Homepage. <https://agricoop.nic.in/en>. Accessed 18 Feb 2022
24. NASA Prediction Of Worldwide Energy Resources (POWER). <https://power.larc.nasa.gov/>. Accessed 01 Mar 2022