FarmRight - A Crop Recommendation System

Sakshi[*], Sanjana Drall[†], Sukriti Singh[‡], Dviti Arora[§], Ms. Monica Choudhary [¶]

May 7, 2022

Abstract

Agriculture is extremely vital to our economy. Boosting the development of this sector has always been a priority for government.Health of all types of crops that are grown around us is affected by various types of factors which includes temperature , rainfall pattern , type of soil , pH of soil and these factors have been drastically changing since last some decades posing challenges in front of farmers.They confront a huge challenge in selecting the best crop for their property to maximize yield and profit given the changing factors.Farmers sometimes make mistakes when it comes to selecting the appropriate crop and sowing area for production depending on environmental conditions. There is no trustworthy agriculture recommendation system for Indian farmers. Investments, money, and labour are all squandered if the wrong crop or planting place is chosen. Farmers are scared to seed huge regions these days because of unexpected yields, unpredictable climate fluctuations, and a lack of assets. To address this problem, this research presents a crop recommendation system based on a multi-label classification model for crop recommendation that takes into account the user's state, district, soil qualities, and climate features to generate a list of recommended crops. Researchers examine several algorithms based on their performance criteria to find the most accurate crop recommendation model. The most accurate model was discovered to be an RF Technique, with a precision of 82.74 percent, a recall of 80.92 percent, and an F1 score of 78.67 percent. The model that has been trained thus far can be utilized to provide the user with a list of recommended crops. Keywords: crop recommendation, multi-labels classification, random forest algorithm, soil characteristics, climate properties, agriculture.

1 Introduction

Agriculture is extremely vital to our economy acting as the lifeblood of rural households with 70%[1] of them depending on agriculture for their livelihoods. Agriculture is a significant part of the country's growth. It accounts for about 13.7% [MK21]of India's total GDP. It is the largest employment sector that employs more than 57% [MK21]of the population which makes least two-thirds of the working population in India. In other industries, India has not been able to generate many job possibilities for the rising workforce.The rank of our country is second in the agricultural sector of the world. It also impacts international trade and food supply by meeting all the country's needs for grains.Not only this ,production from the agricultural sector serves as an input to the industrial sector. As more and more businesses are established and existing businesses expand, the demand for raw materials increases. This increasing demand can only be met by increasing agricultural production. So , there is increasing pressure on the country's farmers to meet the demand for plants. With that in mind, we focus on dramatic changing factors such as weather and soil that affect plant health. Today's farm owners face significant difficulty in selecting the optimum crop for their land to boost yield and profit. They need to choose the best harvest for their land. They have picked favorable features in crops for thousands of years, improving the plants for agricultural reasons. They started looking for crop types with shorter growing seasons, larger seeds, and fruits. Those crops must have higher disease and insect resistance, shelf life, nutritional content, and better adaptation to varied ecological

[*]Computer Science and Engineering IGDTUW, sakshi097btcse18@igdtuw.ac.in
[†]Computer Science and Engineering IGDTUW, sanjana107btcse18@igdtuw.ac.in
[‡]Computer Science and Engineering IGDTUW, sukriti110btcse18@igdtuw.ac.in
[§]Computer Science and Engineering IGDTUW, dvitiarora125btcse18@igdtuw.ac.in
[¶]Assistant Professor, CSE Department, IGDTUW , monikachoudhary@igdtuw.ac.in
[1]https://www.fao.org/india/fao-in-india/india-at-a-glance/en/

environments where crops were cultivated were all desirable qualities. Agricultural technology has produced a wide range of possibilities for food, feed, and fiber production over the years. In many respects, technology makes our lives easier and more enjoyable by reducing the amount of time we devote to essential operations like food production. Agriculture has seen significant improvements, many of which improve the efficiency and safety of food and textile production. Sometimes farmers do not choose the right crops and suitable seeding areas to plant-based on environmental factors. There is no reliable agricultural recommendation system for Indian farmers. 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 [KS22] 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 [EV21] .The recommendation system uses multi-label classification which is a supervised learning task where an instance may be associated with one or more than one mutually non-exclusive class labels.The comparison of different algorithms based on their performance metrics is used to find the most accurate model for crop recommendation. The most accurate model was found to be a Random Forest Algorithm with a precision of 82.74%, recall 80.92%, and an F1 score of 78.67%. The model trained hence, can be used for giving a list of the recommended crops to the user.

The rest of this paper is organized as follows: In section 2, we provide a brief review towards some related work on crop recommendation systems.Section 3 contains the details about the solution that paper aims at proposing , followed by Section 4 which explains the steps followed for making the model.Experimental results are presented in section 5. Finally Section 6 concludes the paper and discusses about the future scope.

2 Background and Literature Review

Given the importance of agriculture for our economy , researchers have been trying to devise 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. [VSK17] develops a project to determine soil fertility and proposes that crops 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. Krupa et al. [PP20] 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 based on different parameters. S Pudumalar et al. [PRR+17] 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, an advanced cultivation technique that uses research data on soil characteristics, type of soil, and yield data to advise cultivators on the optimum crop for their particular site has been used as the base for this research work.This solution is a system of recommendation that uses algorithms such as CHAID, KNN, Naive Bayes etc as learners.Banerjee et al. [BSG21] 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. [LLWC20] 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. [KSKS15] 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. [SP13] proposed SVM model for forecasting exports of agricultural products and used Autoregressive Integrate Moving Average (ARIMA) as a benchmarking to compare other developed models and used data of Thailand's Pacific white shrimp export and Thailand's produced chicken. Ramya et al. [RBG15] aims at providing a Climate Smart and agriculture decision-making platform for changing climate that affects agriculture. Using data and analytics, they predicted and mitigated the impact of extreme weather events on global finance, as well as the Economic Dimensions of Climate Change around the world. Kulkarni et al. [KSSC18] 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.As Machine Learning algorithms advanced , Jeong et al. [JRM+16] 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. Krishna et al. [KKRKA+04] studies and presents an analysis of crop–climate relationships for India, using historic production statistics for major crops (rice, wheat, sorghum, groundnut and sugarcane) and for aggregate food grain, cereal, pulses and oilseed 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 [KSSC18] though there are more than hundreds of crops that are grown in India.Moreover, dataset used focused on data from few locations as in [PRR+17] and [VSK17].Algorithms used and the accuracy of the model were not mentioned clearly as in [VSK17].

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.

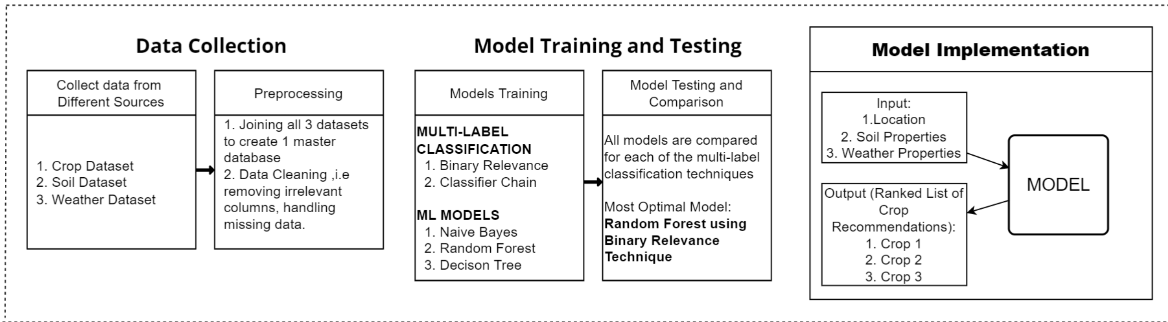

Figure 1: Workflow Diagram

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. The three major elements are nitrogen(N), phosphorus(P), and potassium(K) which are required in high amounts. Soil pH is a major variable in soils as it controls many chemical and biochemical processes occurring within the soil. It is the degree of acidity or basicity/alkalinity of the soil. Soil pH study is critical in agriculture because it governs plant nutrient availability by managing the chemical forms of various nutrients as well as influencing their chemical interactions. As a result, the pH value of the soil has a strong influence on soil and crop productivity. India's farmers rely on the monsoon, which comes from the Indian and Arabian Seas. Farmers are unprepared for changes in the rainfall cycle, magnitude, and timing as the climate changes. Water is retained in the form of moisture when the temperature is warm. 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:

i) Crop Dataset[2]: 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 [3].

ii) Soil Dataset[4]: This dataset sourced by the DoAFW [5] 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.

iii) 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 Meters (°C) ,Precipitation Corrected (mm).The dataset had around 14212 instances. This dataset was collected from API from NASA Prediction Of Worldwide Energy Resources (POWER)[6].

[2] https://github.com/Know-and-Grow/FarmRight-A-Crop-Recommendation-System/blob/main/Datasets/Crop$_{D}B.csv$
[3] https://data.gov.in/
[4] https://github.com/Know-and-Grow/FarmRight-A-Crop-Recommendation-System/blob/main/Datasets/Soil$_{D}B.csv$
[5] https://agricoop.nic.in/en
[6] https://power.larc.nasa.gov/

CROP DATA SET			CLIMATE		SOIL PROPERTIES							LABEL
State_Name	District	Season	Temp	Rainfall	N	P	K	OC	Acidic	Neutral	Basic	Crop
ANDAMA	NICOBAR	Kharif	27.5725	9.145	99.11	78.16	89.89	0.6596	35.27	8.53	56.2	Arecanut
ANDAMA	NICOBAR	Kharif	27.5725	9.145	99.11	78.16	89.89	0.6596	35.27	8.53	56.2	Other Kharif pulse
ANDAMA	NICOBAR	Kharif	27.5725	9.145	99.11	78.16	89.89	0.6596	35.27	8.53	56.2	Rice
ANDAMA	NICOBAR	Whole Ye	27.5725	6.08	99.11	78.16	89.89	0.6596	35.27	8.53	56.2	Banana
ANDAMA	NICOBAR	Whole Ye	27.57	6.08	99.11	78.16	89.89	0.6596	35.27	8.53	56.2	Cashewnut
ANDAMA	NICOBAR	Whole Ye	27.57	6.08	99.11	78.16	89.89	0.6596	35.27	8.53	56.2	Coconut
ANDAMA	NICOBAR	Whole Ye	27.57	6.08	99.11	78.16	89.89	0.6596	35.27	8.53	56.2	Dry ginger
ANDAMA	NICOBAR	Whole Ye	27.57	6.08	99.11	78.16	89.89	0.6596	35.27	8.53	56.2	Sugarcane
ANDAMA	NICOBAR	Whole Ye	27.57	6.08	99.11	78.16	89.89	0.6596	35.27	8.53	56.2	Sweet potato
ANDAMA	NICOBAR	Whole Ye	27.57	6.08	99.11	78.16	89.89	0.6596	35.27	8.53	56.2	Tapioca

Figure 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

Classification is a predictive modeling problem where a class label is obtained as output when an input is given. This differs from regression problems involving the prediction of a numeric value. Classification tasks generally predict a single label. Alternatively, it may involve predicting the probability for two or more class labels. In these cases, the classes are mutually exclusive, i.e the classification task assumes that the input belongs to only one class. However, there are some scenarios where there is a requirement of predicting more than one class label. This means that class labels are mutually non-exclusive. These tasks are known as multiple label classification or multi-label classification for short. Multi-label classification results in zero or more labels as output for each input sample, and the outputs are required simultaneously. The output labels are assumed to be 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. Most traditional learning algorithms are developed for single-label classification problems. Therefore a lot of approaches in the literature transform the multi-label problem into multiple single-label problems at the existing single-label algorithms can be used. Here, we have considered two techniques of multi-label classification:

- Binary Relevance: In this case, for each class, an ensemble of single-label binary classifiers is trained. Each classifier predicts whether an instance belongs to a particular class or not. All classes that were predicted are considered together to give the multi-label output. This is a popular approach because it is very simple to implement, but it doesn't take the possible correlations between class labels into account.

- Classifier Chains: In this case, a chain of n binary classifiers, where n is the number of possible class labels, is constructed, where a classifier uses the predictions of all its former classifiers. In this way this allows us to take label correlations into account.

4.4 Models Used

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

- Naive Bayes: The Naive Bayes [2] classifier, a simple heuristic classifier whose work is based on applying Bayes' theorem (from Bayesian statistics) with the strong naive independent hypothesis. In this technique, classifier models are constructed to assign class labels to instances, These instances are represented as vectors of feature values, and the class labels are drawn from some finite set. It is not just a single algorithm but a family of algorithms based on a common principle. All naive Bayes classifiers assume that the value of one attribute is independent of the value of any other attribute, given the class variable. Naive Bayes is useful for multidimentional information and because of this it is considered in this case. For our system, we have considered two Naive Bayes classifiers, namely Multinomial Naive Bayes and Gaussian Naive Bayes.

- Decision Tree: Decision trees obtain information in the form of a tree structure. This tree structure can also be written as a set of discrete rules for better clarity and easier understanding. The decision tree classifiers have an upper hand in their ability to use different subsets of attributes and decision rules at different stages of classification. A typical decision tree consists of a root node, some internal and leaf nodes, and branches. The class to be assigned to a sample is represented by the leaf node. Each internal node of a tree corresponds to an attribute, and branches denote conjunctions of features that lead to those classifications. Decision tree splits population of data into smaller segments and this makes it better for better predictions.

- Random Forest: RFA is a bagging technique based on the tree ensemble method in machine learning. It creates a series of trees with randomly subsampled features. The average value of the prediction of individual trees is evaluated as the output of RFA. Since it is using random sub-sampled features, RFA can be used in high dimension input predictors.

4.5 Model Training and Performance Metrices

The dataset is 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: The capacity of a classification model to identify only the relevant data points. It can be defined as the number of true positives divided by the number of true positives plus the number of false positives.

- Recall: A classification model's ability to locate all relevant cases within a data collection. The number of true positives divided by the number of true positives plus the number of false negatives is known as 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.

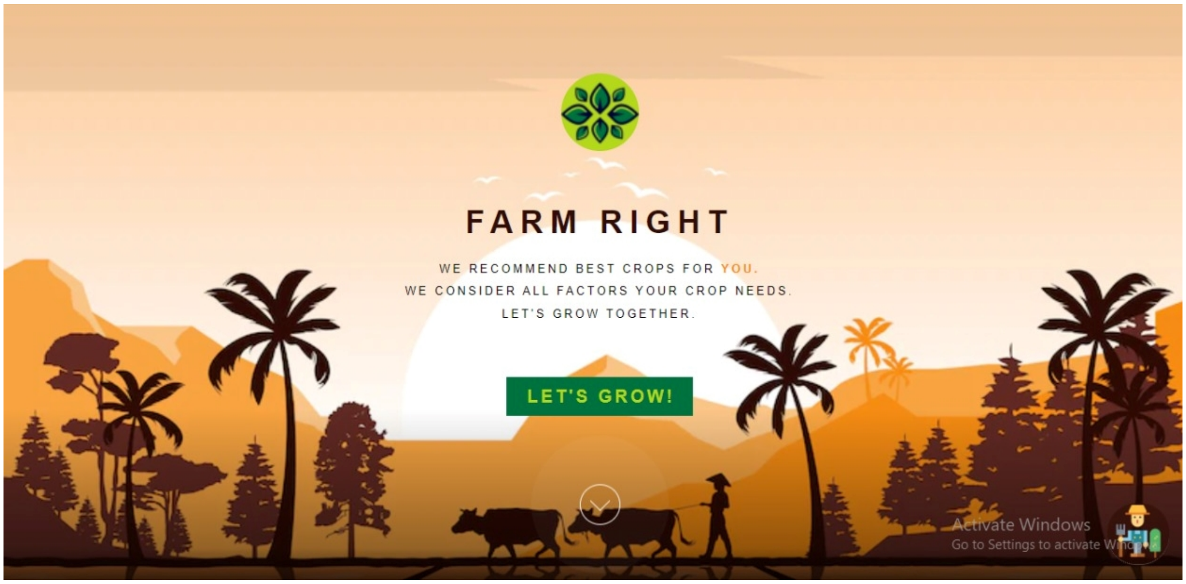

Figure 3: Website Integrated with FarmRight.

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. On the other hand, the RFA using the Binary Relevance technique returned the least hamming loss and high values of precision, recall, and F1 score. Hence, it can be concluded that the RFA with Binary Relevance is the most suited algorithm for our proposed solution.

Table 1: 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: 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 is further integrated with a website [7] to provide an interface to the users for its best use.Any user can visit the website and get to know which is the most suitable crop for his area.The website also has a chatbot which enables users to explicitly ask their queries and get the best answers.

[7]https://github.com/Know-and-Grow/FarmRight-A-Crop-Recommendation-System

6 Conclusion and Future Scope

India is a country where agriculture plays an important role. The country prospers in the prosperity of the peasants. Therefore, our work helps farmers sow the right seeds based on soil requirements and climatic characteristics to increase productivity and utilize techniques like these for their benefit. 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. The comparison of different algorithms based on their performance metrics is used to find the most accurate model. The most accurate model was found to be a Random Forest Algorithm with a precision of 82.74%, a recall of 80.92%, and an F1 score of 78.67%. The model trained hence, can be used for giving a list of the recommended crops to the user.

In future, new and updated dataset could be used for prediction of crops.We can further look for other model building techniques and model tuning can be performed as well, Also, the website could be made multi-lingual, chatbot that can handle extensive queries could be developed for easy user experience, connecting experts and adding helpline numbers could be done for scalable use.Along with this , fertilizer recommendation depending upon region and crop chosen by the user can be added as another feature in the website.

References

[BSG21] 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*, pages 57–69. Springer, 2021.

[EV21] Dhivya Elavarasan and PM Vincent. A reinforced random forest model for enhanced crop yield prediction by integrating agrarian parameters. *Journal of Ambient Intelligence and Humanized Computing*, 12(11):10009–10022, 2021.

[JRM+16] Jig Han Jeong, Jonathan P Resop, Nathaniel D Mueller, David H Fleisher, Kyungdahm Yun, Ethan E Butler, Dennis J Timlin, Kyo-Moon Shim, James S Gerber, Vangimalla R Reddy, et al. Random forests for global and regional crop yield predictions. *PLoS One*, 11(6):e0156571, 2016.

[KKRKA+04] Kanikicharla Krishna Kumar, Kolli Rupa Kumar, RG Ashrit, NR Deshpande, and James William Hansen. Climate impacts on indian agriculture. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 24(11):1375–1393, 2004.

[KS22] Raj Kumar and Vivek Singhal. 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.

[KSKS15] Rakesh Kumar, MP Singh, Prabhat Kumar, and JP Singh. 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)*, pages 138–145. IEEE, 2015.

[KSSC18] Nidhi H Kulkarni, GN Srinivasan, BM Sagar, and NK Cauvery. 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)*, pages 114–119. IEEE, 2018.

[LLWC20] Aoqi Liu, Tao Lu, Bufan Wang, and Chong Chen. Crop recommendation via clustering center optimized algorithm for imbalanced soil data. In *2020 5th International Conference on Control, Robotics and Cybernetics (CRC)*, pages 31–35. IEEE, 2020.

[MK21] Arvind Malik and Ramarcha Kumar. An overview on agriculture in india. *International Journal of Modern Agriculture*, 10(2):2087–2095, 2021.

[PP20] Krupa Patel and Hiren B Patel. A state-of-the-art survey on recommendation system and prospective extensions. *Computers and Electronics in Agriculture*, 178:105779, 2020.

[PRR+17] S Pudumalar, E Ramanujam, R Harine Rajashree, C Kavya, T Kiruthika, and J Nisha. Crop recommendation system for precision agriculture. In *2016 Eighth International Conference on Advanced Computing (ICoAC)*, pages 32–36. IEEE, 2017.

[RBG15] MGPG Ramya, Chetan Balaji, and L Girish. Environment change prediction to adapt climate-smart agriculture using big data analytics. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 4(5), 2015.

[SP13] Thoranin Sujjaviriyasup and Komkrit Pitiruek. Agricultural product forecasting using machine learning approach. *Int. Journal of Math. Analysis*, 7(38):1869–1875, 2013.

[VSK17] PS Vijayabaskar, R Sreemathi, and E Keertanaa. Crop prediction using predictive analytics. In *2017 International Conference on Computation of Power, Energy Information and Commuincation (ICCPEIC)*, pages 370–373. IEEE, 2017.