International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 05 | May 2022 www.irjet.net p-ISSN: 2395-0072

Crop Recommendation System Using Machine Learning

Ajay Lokhande[1], Prof. Manish Dixit[2]

[1]B. Tech Student, Dept. of Computer Science and Engineering, Madhav Institute of Technology and Science, Gwalior, Madhya Pradesh, India
[2]Professor and Faculty Mentor, Dept. of Computer Science and Engineering, Madhav Institute of Technology and Science, Gwalior, Madhya Pradesh, India

---

Abstract - *In India agriculture holds an incredibly predominant position in the expansion of our country's financial system. It is one of the field which generates most of the employment opportunity in our country. Farmers, due to lack of their knowledge about different soil contents and environment conditions, do not opt the exact crop for nurturing, which results into a major hinder in crop production. To eliminate this barrier, we have provided a system which offers a scientific approach to assist farmers in predicting the ample crops to be cultivated based on different parameters which affects the overall production. It also suggests them about several deficiencies of nutrients in the soil to produce a specific crop. It is in context of a website. We used the crop dataset which include parameters like temperature, rainfall, pH, and humidity for specific crops and applied different ML techniques to recommend crops with high accuracy and efficiency. Hence, it can be supportive for farmers to be furthermore extra versatile.*

Key Words: **Crop recommendation, Machine Learning, Random Forest, SVM, Logistic Regression.**

1.INTRODUCTION

Foreign nations have started using and implementing modern methods for the purpose of their profit. They already have gained such an upper hand in applying scientific and technological methods in the field of agriculture and farming to increase its quality of work that one can only imagine. Whereas, India still is holding on to the traditional approach towards farming and its technologies. As, we know, that singularly agriculture alone generates a huge percentage of revenue for our country. The income come is pretty handy, when we talk about the Gross Domestic Product value. Moving forward towards globalization, the need for food has grown exponentially. Farmer tries to increase the quantity of their production by adding different artificial fertilizers, which eventually results in a future environmental harm. But if the farmer knows exactly which crop to be sown according to different soil contents and environment conditions, then this will minimize the loss and results in efficient crop production. We have gathered a dataset, which consist of information about rainfall, climatic conditions and different soil nutrients. This will provide us better understanding of trends of crop production in consideration of geographical and environment factors. Our system also predicts the shortage of any particular components for growing a specific crop. Our predictive system can prove to be a great boon for agricultural industry. The trouble of nutrient insufficiency in areas, which occurred due to the reason of planting incorrect crop at wrong period of time, is abolished with the help of our predictive system. This results in scaling down the production efficiency of farmers. More and more scientific approach towards agriculture industry, will definitely take it to greater heights.

We propose this system to provide farmers, knowledge about requirements of different minerals and climatic conditions, which are suitable for producing certain crops. Also, our project diverts our focus towards the lack of different minerals required to grow some crops, and proposes us the remedies to eliminate their shortage. Our system considers factors such as soil composition and climatic factors like temperature, rainfall and humidity.

2. Literary Survey

[3] Crop Recommendation System for Precision Agriculture - S.Pudumalar*, E.Ramanujam*, R.Harine Rajashree, C.Kavya, T.Kiruthika, J.Nisha. The system developed is based upon a combination of specific ML techniques. It uses the soil nutrient data, yield data and soil type data in order to predict the suitable crops for harvesting. The model devised recommends us the most likely result based on the input entered. It is an ensemble model technique making use of algorithms such as Random tree, CHAID and SVM.

[4] Prediction of Crop Cultivation – Rikhsit K. Solanki, D Bein, J. A. Vasko & N. Rale. – A lot of ML technologies is being put under the test. The basic objective that the paper offers is to minimize the incorrect crop selection risks by making a system with optimum prediction accuracy rate. Here, the author uses k- fold cross validation techniques, where k is taken to be 5. This usually means that it is a 5 – fold cross validation, where we split given data into 5 sets, where about (k-1) i.e. 4 sets of data is used for training and the remaining 1 set is used for testing process. Authors of the article, simplified their research and concluded that with the best performance attribute results, the RF machine learning

technique leads the chart, followed by support-vector regression using RBF kernel.

[5] Supervised Machine learning Approach for Crop Yield Prediction in Agricultural Sector – Dr. J. N. Kumar, V. Spandana, V.S. Vaishnavi – In this system depiction, the researchers found the results from old farming data and expected yield production accordingly. Also the system supports the prediction of harvest both qualitatively and quantitatively with at most efficiency.

[6] Improving Crop Productivity through a Crop Recommendation System Using Ensembling Technique - N K Cauvery, Nidhi H Kulkarni, Prof. B. Sagar, N. Cauvey. In this system, crop proposal structure is to be developed that make use of the combinational practice of knowledge gaining procedure. This routine is used to create a representation that joins the predictions of abundant ML tactics as one to advise the exact produce based upon the uniqueness with elevated precision.

3. System Design

In order to develop a software program we must be familiar with concept of SDLC life cycles. A SDLC cycle is a software development framework which allows a user to manage and implement its system in an efficient manner. In our project the model that we have used is the waterfall model. . It is a very basic and easy to use life cycle model. It uses a linear sequential flow to make the software. This means that the next process cannot be started without completing the previous process. Thus, there is no overlapping of process. The outcome of one segment is the input for the then part. The waterfall model includes the stages as shown in the "Fig.1". All of the below mentioned stages are very important for system development and input of a process is the outcome of previous one itself.

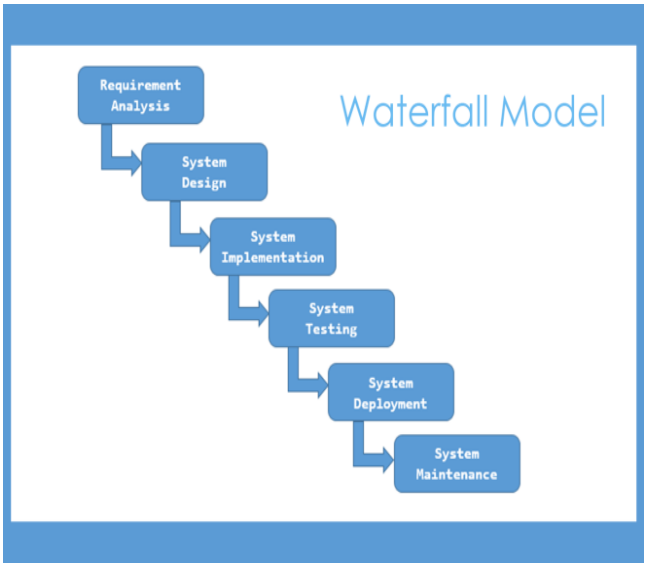


Fig. -1: Waterfall Model

3.1 Model Framework Overview

The most important steps involved in building a machine learning based predictive model are shown in the following " Fig. 2".

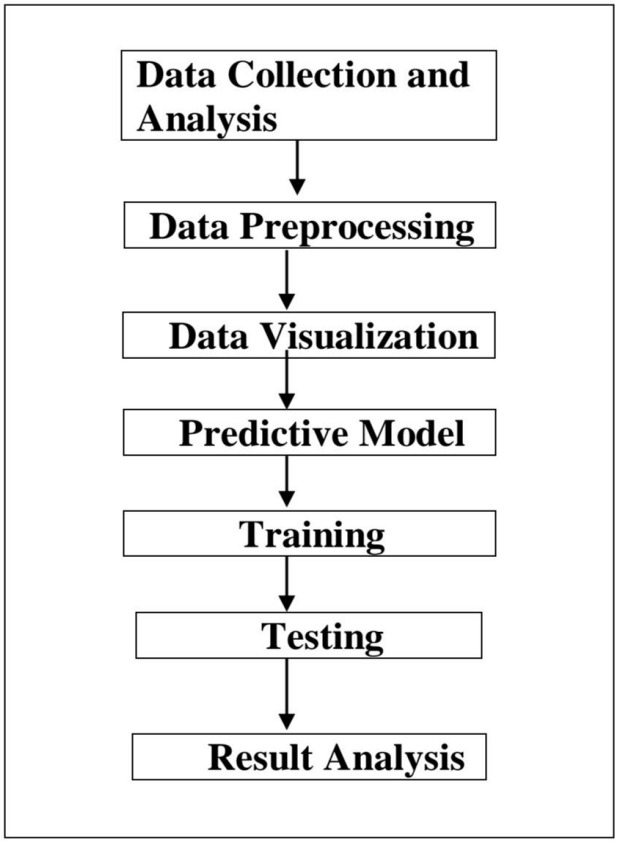


Fig. -2: Project Framework

3.2 System Architecture

A system architecture is a conceptualization oriented model that is used to define the structure and behavior of our project. An architecture is a formally descriptive and representative structure of a system which is organized in a way that supports the logic behind every structures and behavior of the system. It also represents the relationship or dependencies of one process to another. A good system architecture helps us to understand and learn about a system in a graphical manner. . It allows us to learn about our product in such a way that we can easily relate and understand the internal processes, methods and stages of a system. The system architecture of our recommendation model is depicted in the "Fig. 3".

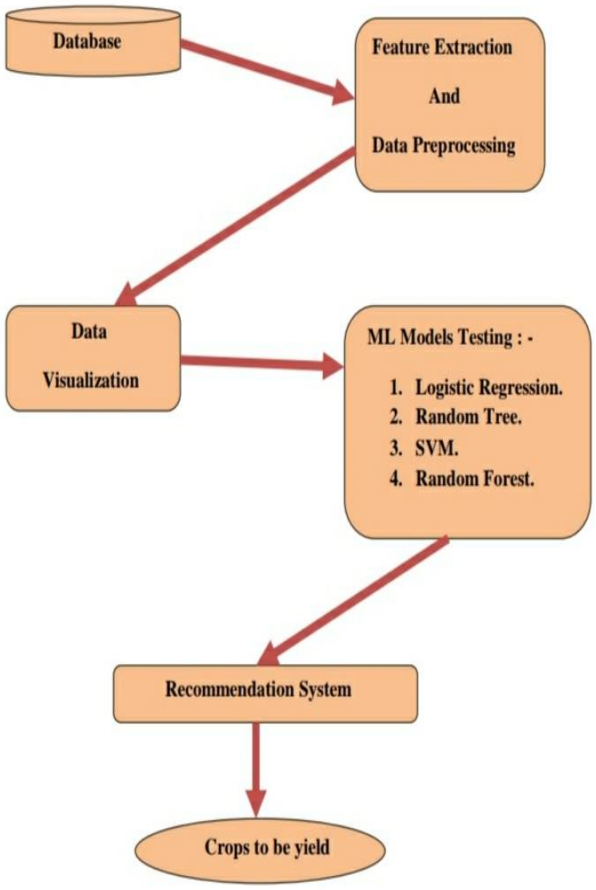


Fig. -3: System Architecture

4. Methodology and Implementation

The methods and the processes to be carried out for implementation of our predictive system is stated below in a sequential manner.

4.1 Dataset Collection

The first step that we perform during machine learning project development is the collection of dataset. The dataset we obtain from various platforms is the raw data having a tremendous amount of errors and ambiguities in it.

In this project we have obtained our data from an open source platform known to as kaggle. Our dataset is basically a collection of two more intricate datasets. The two datasets are : soil content dataset (consisting information about ratios of Nitrogen(N), Phosphorous(P), and Potassium(K) and ph of the soil) and climatic condition dataset (containing information about rainfall, humidity and temperature).The final dataset comprises about of about 2200 rows and about 8 columns. “Fig. 4” depicts our final dataset.

	A	B	C	D	E	F	G	H
1	N	P	K	temperature	humidity	ph	rainfall	label
2	90	42	43	20.87974371	82.0027442	6.5029853	202.9355362	rice
3	85	58	41	21.77046169	80.3196441	7.0380964	226.6555374	rice
4	60	55	44	23.00445915	82.3207629	7.8402071	263.9642476	rice
5	74	35	40	26.49109635	80.1583626	6.9804009	242.8640342	rice
6	78	42	42	20.13017482	81.6048729	7.6284729	262.7173405	rice
7	69	37	42	23.05804872	83.3701177	7.0734535	251.0549998	rice
8	69	55	38	22.70883798	82.6394139	5.7008057	271.3248604	rice
9	94	53	40	20.27774362	82.8940862	5.7186272	241.9741949	rice
10	89	54	38	24.51588066	83.5352163	6.6853464	230.4462359	rice
11	68	58	38	23.22397386	83.0332269	6.3362535	221.2091958	rice
12	91	53	40	26.52723513	81.4175385	5.3861678	264.6148697	rice
13	90	46	42	23.97898217	81.450616	7.502834	250.0832336	rice
14	78	58	44	26.80079604	80.8868482	5.1086818	284.4364567	rice
15	93	56	36	24.01497622	82.0568718	6.9843537	185.2773389	rice
16	94	50	37	25.66585205	80.6638505	6.9480198	209.5869708	rice
17	60	48	39	24.28209415	80.3002559	7.0422991	231.0863347	rice
18	85	38	41	21.58711777	82.7883708	6.2490507	276.6552459	rice
19	91	35	39	23.79391957	80.4181796	6.9708598	206.2611855	rice
20	77	38	36	21.8652524	80.1923008	5.9539333	224.5550169	rice
21	88	35	40	23.57943626	83.5876032	5.8539321	291.2986618	rice
22	89	45	36	21.32504158	80.474764	6.4424754	185.4974732	rice
23	76	40	43	25.15745531	83.1171348	5.0701757	231.3843163	rice
24	67	59	41	21.94766735	80.973842	6.0126326	213.3560921	rice
25	83	41	43	21.0525355	82.6783952	6.2540285	233.1075816	rice

Crop_recommendation

Fig. -4: Final Crop Dataset

4.2 Data Analysis

This step probably comes before the preprocessing step. In this step we try to understand the data with more concentrating. In this step we basically try to extract some features from data which will be taken as a parameter to predict the results. This step should be handled with at most patience. As, on its basis only we will be predicting our crop results. We have provided a heatmap which represents the data analysis process of our recommender system in “Fig. 5”. This step probably comes before the preprocessing step. In this step we try to understand the data with more concentrating. In this step we basically try to extract some features from data which will be taken as a parameter to predict the results. This step should be handled with at most patience. As, on its basis only we will be predicting our crop results. We have provided a heatmap which represents the data analysis process of our recommender system in “Fig. 5”.

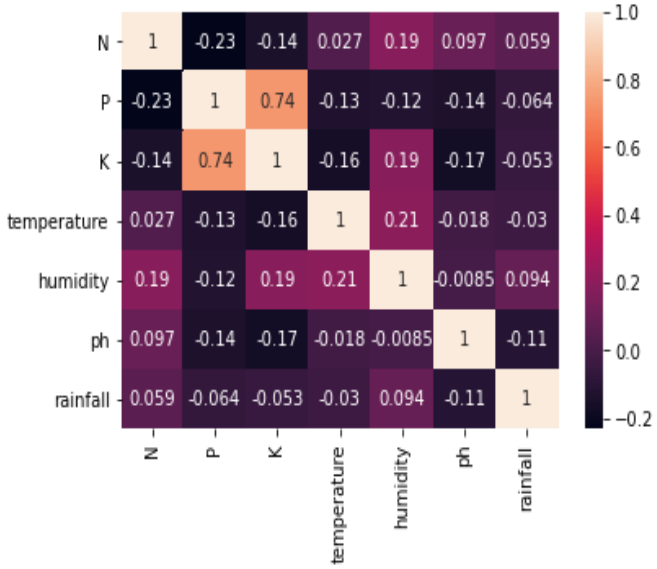


Fig. -5: Data Analysis as Heatmap

5. Data Visualization

We have plotted the production quantity of different cops versus the factors that affect its production. This will provide us with knowledge of the how production quantity is affected by various climatic factors as well as soil components. The data we have plotted is in the form of bar graphs can be seen in the "Fig.6".

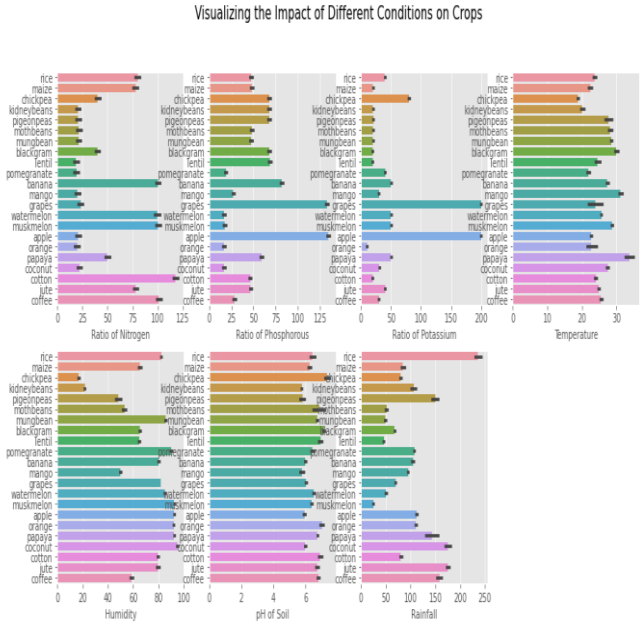


Fig. -6: Data Visualization

6. Machine Learning Algorithms

Since our crop prediction system is a type of classification problem, we trained and tested numerous algorithms. To get optimum accuracy and the most suitable crop to be sown, we made sure that its accuracy is prioritized. We compared the accuracy of all the models and choose who has highest of all.

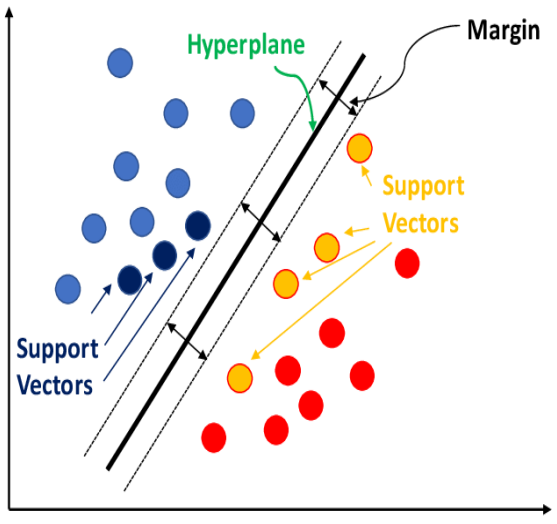
6.1 Decision Tree

The first ML method that we made use of in order to build our target crop system is the Decision Tree. The fact is that, it gained its name, because of the graphical representation and its parsing structure that it follows. It consists mainly of two types of nodes: decision and the leaf nodes. Decision nodes that has branching and carry out the functionality of decision making and the leaf nodes are the nodes with no branching and displays final outcomes of the decisions with given conditions.

The accuracy of the tested model comes out to be about 90.0%. Thus, its accuracy is not too bad, but we have not taken it because of the precision is not so great.

6.2 Support Vector Machine

SVM is a method of ML that generates an optimal hyperplane or decision boundary that can segregate dimensional spaces into classes, for the purpose of putting the new data in correct category in future. With the help of support vectors, we create hyperplane. The hyperplane thus generated has two support vectors each on either side of the hyperplane. The support vectors are nothing but lines that is drawn passing through two data points on either side, which is closest to hyperplane.
The accuracy of this model is about 96.08 %. Thus, this model turns out to be more accurate than Random tree algorithm. The "Fig.7" is depiction of SVM classifier.



Fig, -7: SVM Classifier

6.3 Logical Regression

Logical regression is the one of the ML algorithm whose approach is to model bond between dependent along with independent variables. It is used to solve categorical data

problems mainly. It is straightforward to implement and very fundamental and efficient model.

The accuracy of logical regression model is 95.22 %. Its accuracy is better than random tree algorithm, but is lower than SVM. Thus, this model is discarded.

6.4 Random Forest

We already have used, learned and tested how decision tree method works. So, understanding random forest method will not be so difficult. It is a very popular algorithm based on ensemble learning. It is a amalgamation of a bunch of random trees on different subnets of the dataset. Presence of excessive random trees in a random forest network, leads to high accuracy of model. It takes the decision from the tree with most votes. Hence, more the number of decision trees in the forest, greater the accuracy of the model. This also eliminates the problem of overfitting.

The accuracy of the random forest algorithm is highest with 99.09 %. Thus at last we take this algorithm to use in our model building and final implementation of our project as a website. The below "Fig. 8" is a representation of random forest model. In general manner, we say that forest is a combination of trees. Here also we can say that, random forest is a combination of various subnets of decision tree.

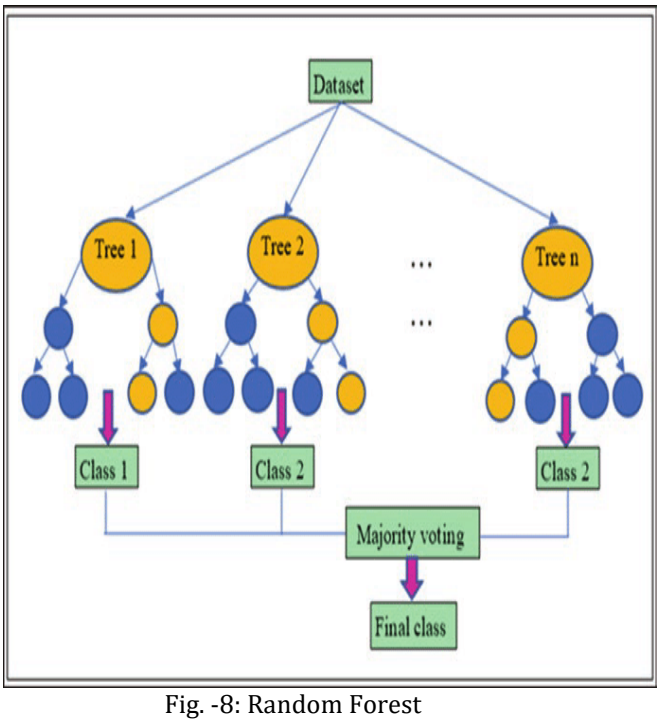


Fig. -8: Random Forest

The random forest model is than converted into pickle file. This pickle file is then used with flask, html, css and some basic javascript language in order to implement and run as a website on our local host.

7. Testing

We have used here cross validation training method is a validating technique that is used to verify how the arithmetic model generalizes to a sovereign datasets. Here, we have used K- folding cross validation. In this type of cross validation we only use about 1 set of data to train, and other else are used for training our mold.

Also our crop recommendation system project is a combination of all the functional units of the program developed separately. Several processes were conducted as mentioned in system architecture. Here, in this project we have tried and tested several machine learning algorithms in order to obtain a good and accurate model for farmers to predict crops. All the models are tested for the attributes such as accuracy, precision and FP rate and others. We have also checked cross validation score for each and every algorithm that we tried to built this system.

Our recommendation method replica is tested for different performance attributes. Each of the performance attribute is explained below.

7.1 Accuracy

Accuracy is the most prominent factor for the evaluation of a machine learning model. On the basis of accuracy, we decide whether our machine learning model is applicable in real world or not. If the accuracy of an algorithm implemented is high, then eventually it means the system is more closer to real world.

$$\text{Acc.} = \frac{\text{No. of Correct Predictions}}{\text{Total No. of Predictions}}$$

Also, it is evaluated as:-

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

7.2 Precision

Precision is calculated by simply plotting the confusion matrix. We can find the TP, FP, TN and FN values from it. In order to evaluate precision we take True Positive value in numerator and the sum of True positive and False Positive. Formula is :-

$$\text{Precision} = \frac{TP}{TP + FP}$$

7.3 Recall

When we take True Positive value in numerator and the sum of True positive and False negative. It shows how much proportion does the actual positive is identical.

$$\text{Recall} = \frac{TP}{TP+FN}$$

8. Result and Performance Analysis

A detailed analysis is carried out on this dataset, and various results were generated in testing phase for checking the accuracy of the model, so that we can devise a system which recommends us optimal crop. Also our system predicts nutrients deficiency. The below "Fig.9" shows results in Jupyter notebook.

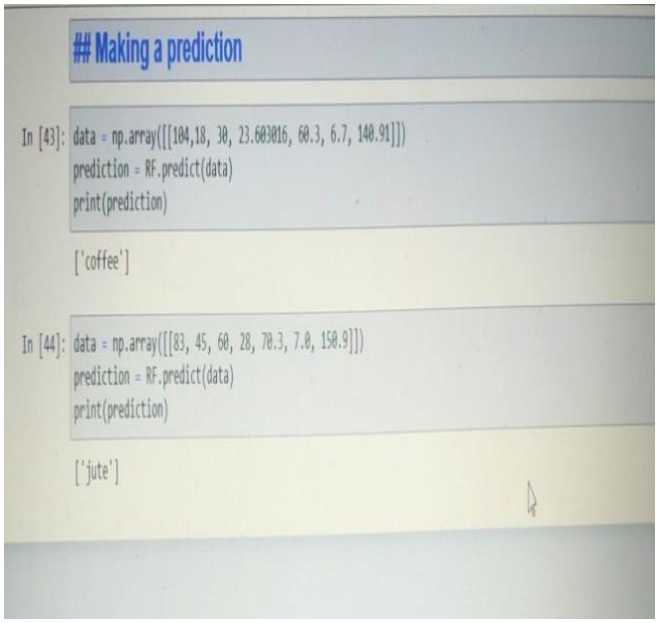


Fig. -9: Results of Prediction in Jupyter Notebook

The data is analyzed and visualized properly to get full understanding of different attributes that affect crops. Different machine learning algorithms are methods that have been tested and checked for accuracy. The correctness proportion of diverse algorithm is plotted and made known in the "Fig. 10".These accuracy percentages is depicted in the form of bar graph.

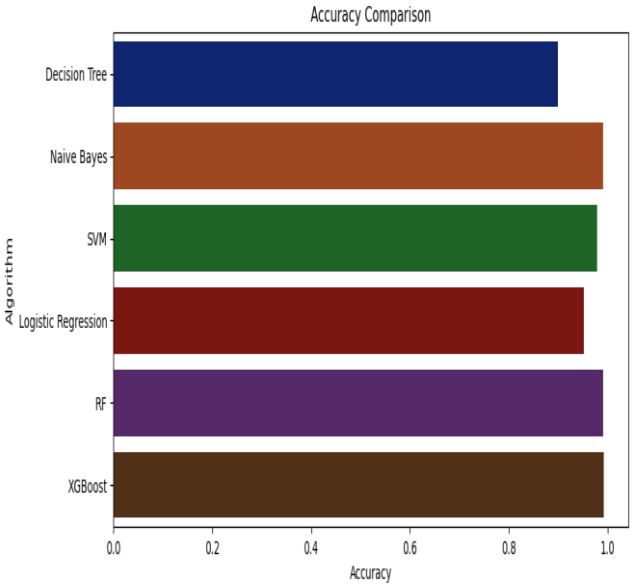


Fig. -10: Accuracy comparision of Algorithms.

We finally came to conclusion to use Random Forest algorithm to Built our predictive system. The accuracy of our Random Forest based machine learning model came out to be about 99.09%. Thus we can say that there is very less margin of error in our prediction. Our model is implemented in a form of a website, which makes use of html,css, flask and javascript. It is deployed on heroku. The interface as an webpage is displayed of our system is shown in "Fig. 11".

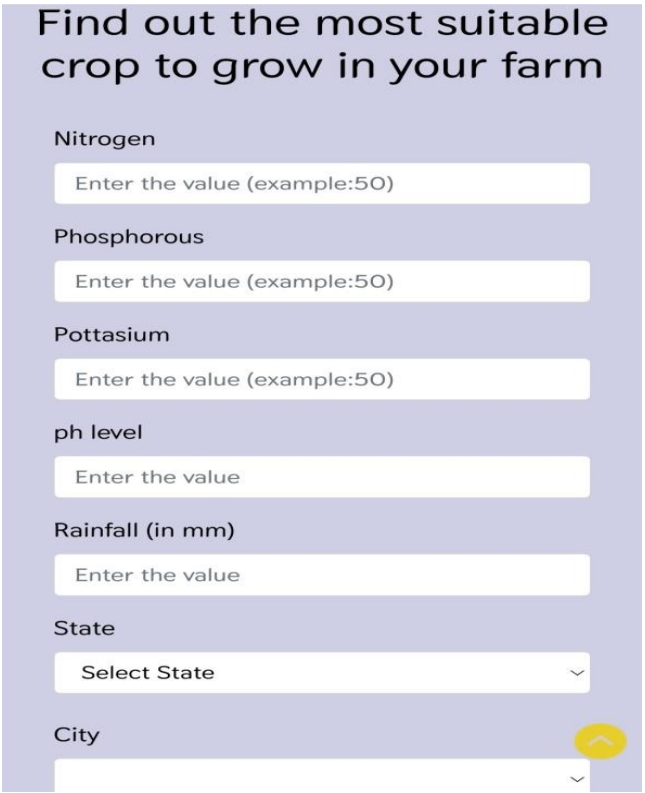


Fig. -11: Interface of Crop Prediction System

Results of our Crop Recommendation System is shown in "Fig. 12". Our system predicts the optimum crop to be cultivated in the user interface.

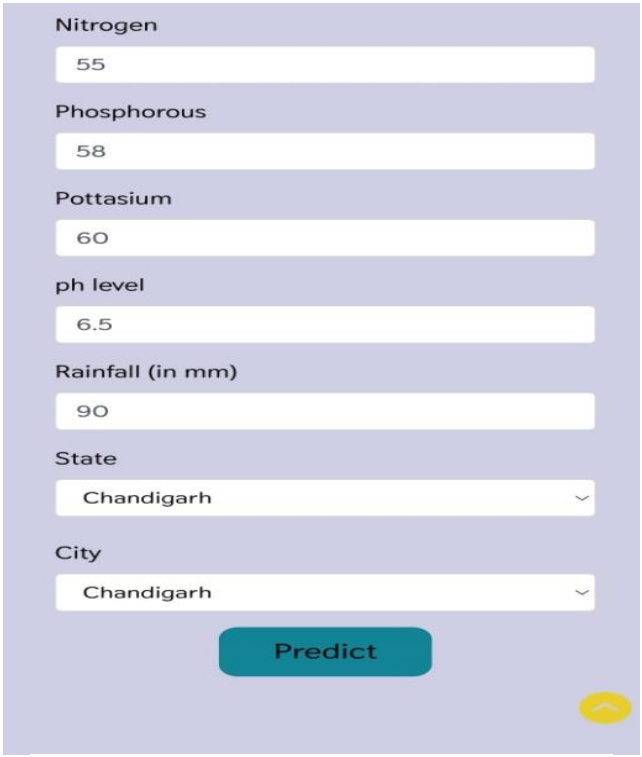


Fig. -12: Result of Predicted Crop

9. CONCLUSIONS

Agriculture is key to our country's economy, so even the smallest investment done in this sector has a tremendous effect on our country altogether. Therefore we need to be more serious about it. Due to the lack of scientific knowledge about different factors affecting crops, the farmers of our country tend to face a lot of challenges in selecting right crops to grow. Hence, face a loss in their profit, due to less productivity. But our system will provide them with a ray of hope to grow crops which will earn them at most profit. The quality as well as quantity of their production will increase exponentially. Also it will also help them in maintaining nutrients content in the soil. Both the quantity and quality will be increased.

10. Future Work

The objective of our system is making a robust model. Also, trying to incorporate bigger dataset. In future I am trying to improve the CSS and UI of my website. I am also trying to implement more features to my project such as Plant disease prediction. Trying to learn and implement web-scrapping technique for data collection.

REFERENCES

[1] https://en.wikipedia.org/wiki/Agriculture_in_India.

[2] https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset

[3] Crop Recommendation System for Precision Agriculture S.Pudumalar*, E.Ramanujam*, R.Harine Rajashree, C.Kavya, T.Kiruthika, J.Nisha. In 8th International Conference on Advanced Computing (ICoAC), IEEE, 2017.

[4] Prediction of Crop Cultivation – Rikhsit K. Solanki, D Bein, J. A. Vasko & N. Rale. In 9th Annual Computing and communication Workshop and Conference (CCWC), IEEE, 2019.

[5] In the 5th International Conference on Communication and Electronics Systems (ICCES), June 2020. Dr. J. N. Kumar, V. Spandana, V.S. Vaishnavi have presented their research on Supervised Machine learning Approach for Crop Yield Prediction in Agricultural Sector.

[6] Improving Crop Productivity Through A Crop Recommendation System Using Ensembling Technique – N. K. Cauvey, Nidhi H. Kulkarni, Prof. B. Sagar, N. In 2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS), pp. 114-119. IEEE, 2018.

[7] Nilesh Dumber, Omkar Chikane, Gitesh Moore, "System for Agriculture Recommendation using Data Mining" , E-ISSN : 2454- 9916 | Volume : 1 | Issue : 4 | NOV 2014.

[8] Tom M. Mitchell, Machine Learning, India Edition 2013, McGraw Hill Education.

[9] David C. Rose, William J. Sutherland, Caroline Parker, "Decision support tools for agriculture: Towards effective design and delivery", Agricultural Systems 149 (2016) 165–174, journal homepage: www.elsevier.com/ locate/agsy.