



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 6, June 2022

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.118

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Precision Farming By Analyzing Soil Moisture and NPK Using Machine Learning

Y.Vineela Devi ¹, T.Akshara ², S.Mohitha ³, V.Venkatesh ⁴, N.Sri Hari ⁵

U.G. Student, Department of Computer Science & Engineering, Venkatadri Institute of Technology, Nambur, Andhra Pradesh, India ¹

U.G. Student, Department of Computer Science & Engineering, Venkatadri Institute of Technology, Nambur, Andhra Pradesh, India ²

U.G. Student, Department of Computer Science & Engineering, Venkatadri Institute of Technology, Nambur, Andhra Pradesh, India ³

U.G. Student, Department of Computer Science & Engineering, Venkatadri Institute of Technology, Nambur, Andhra Pradesh, India ⁴

Associate Professor, Department of Computer Science & Engineering, Venkatadri Institute of Technology, Nambur, Andhra Pradesh, India ⁵

ABSTRACT: Farming is one of the major sectors that influences a country's economic growth. In India, the major occupation of most of the people is agriculture. Many new technologies, such as machine learning are being implemented into agriculture so that it is easier for farmers to grow and maximize their yield. In this project, we present a website in which the following applications are implemented: Crop Recommendation and Fertilizer Recommendation respectively. In the Crop Recommendation application, the user will collect the soil data from their side and the application will predict which crop should the user grow. For the fertilizer recommendation application, the user will input the soil data and the type of the crop they are growing, and the application will analyze the data and give suggestions to increase or decrease nutrients level. The Machine Learning Algorithms used in this project are Random Forest, Support Vector Machine, Naive Bayes. The main aim of this project is to recommend a crop which will give us better yield.

KEYWORDS: Phosphorus, Potassium, Crop Recommendation, Fertilizer Recommendation, Support Vector Machine, Random Forest, Naive Bayes.

I. INTRODUCTION

Precision farming is a booming area. Precision agriculture is an advanced agriculture method that utilizes the data of soil chemistry, soil structures and types, yield of crops, weather conditions and helps the farmers to grow crops which will give good yields and also profitable. This helps the farmers to grow profitable crops and through this we can achieve sustainability. In order to overcome the crisis in agriculture field, there is a requirement of excellent crop recommendation applications and fertilizer recommendation applications to provide farmers a better and clear idea in choosing crop for their land and fertilizer for their crops. For this recommendation system we use Support Vector Machine Decision tree classifier and Gaussian Naive Bayes classifier to classify the crops and recommend them based on the soil and regional characteristics.

II. PROPOSED WORK

The main aim of this project is to recommend crop which will give better yield. The crops data is collected from Kaggle website posted by Siddharth Sharma a year ago gathered over the period by Indian Council Of Agricultural Research. The selected crops are Rice, Maize, Kidneybeans, Blackgram, Lentil, Coconut, Cotton, Jute etc.



There are 3 steps in proposed work.

1) Soil Classification:

Soil will be classified based on the nutrients present in the soil like N, P, K, PH value, Soil moisture, Humidity, Temperature and Rainfall. The crops will be classified based on the above values.

2) Crop Recommendation:

Crop Prediction can be done using crop data, nutrients and location data. These inputs are passed to Random Forest, Naïve Bayes and Support Vector Machine algorithms. These algorithms will predict the crop based on the analysis of the data.

3) Fertilizer Recommendation:

Fertilizer will be recommended by analyzing the soil nutrients and it will take NPK values and Crop as input and the system will recommend the level of nutrients needed for the crop.

III. DESIGN AND METHODOLOGY

The main ideology of the research revolves around the concept of identifying the most suitable crop to be grown with the help of a machine learning model. Thus, the results can prove to be exceedingly beneficial for the agriculture farmers. The data collected for this project contain 2200 records in which 22 different crops present. We have applied machine learning algorithms on the dataset. The dataset is used to train the model according to the actual values and then test the model for its accuracy. Presently our farmers are not using technology and analysis, so there may be a chance of wrong selection of crops for cultivation which will reduce their income. Our research aims to take into consideration the ground reality reflecting the actual requirements instead of assuming any of the factors. Since we are mainly catering to uneducated farmers, we have made it more and more visual and easier to operate. This process is laying the foundation for further evaluation through addition of secondary factors having an impact on deciding the crop to be grown on a particular field in a particular place.

3.1 Research Approach

To work on datasets we used scikit-learn library. The model is trained in such a way that it learns from the data given to it and applies the same trends and knowledge to give optimum results for any given input. The model is then implemented after testing to give a visual look and clarity in accessing the machine learning results from the front-end.

The following table shows the importance of features in the descending order. Rainfall has most importance and ph value has least importance.

	Varname	Imp
6	rainfall	0.241402
4	humidity	0.208684
2	K	0.182873
1	P	0.147355
0	N	0.101607
3	temperature	0.067389
5	ph	0.050689

Figure 1: Feature importance

3.2 Method Approach

The dataset is loaded in the ML model. Then after splitting of data into training and testing sets, it is trained and tested using three different algorithms. The algorithm which is recommending the crop accurately will be selected. This model now performs its action by taking in different factors as input and returning the optimal crop yield.

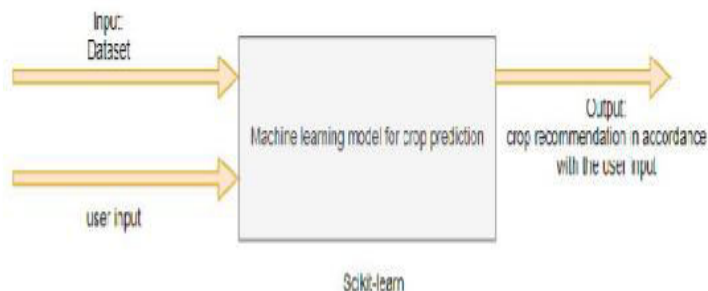


Figure 2: Design of Machine Learning Algorithm

3.3 Kaggle

Kaggle is a platform which contains datasets published by authorized people. Apart from this, it also allows the users to build models in an environment that is generally web-based and data-science oriented. Basically, it is a community for machine learning and data science enthusiasts to get data to work and a platform to display their work. It also hosts competitions where people can compete and hone their ML skills and also get some useful research ideas. We have taken our dataset from Kaggle which we have used to train our model.

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice
...
2195	107	34	32	26.774637	66.413269	6.780064	177.774507	coffee
2196	99	15	27	27.417112	56.636362	6.086922	127.924610	coffee
2197	118	33	30	24.131797	67.225123	6.362608	173.322839	coffee
2198	117	32	34	26.272418	52.127394	6.758793	127.175293	coffee
2199	104	18	30	23.603016	60.396475	6.779833	140.937041	coffee

Figure 3: Crop Recommendation Dataset

3.4 Machine Learning

We have used machine learning to make our model capable of suggesting the optimum crop which can be sown by a farmer according to various input factors. Machine learning is a method of analyzing data to automate the building of an analytical model. It is in fact a branch of AI as it is based on the concept of systems learning from data and identifying some patterns to make decisions without much human intercession.

IV. ALGORITHMS

4.1 Support Vector Machines (SVM)

Support Vector Machines or SVM consists of a group of algorithms that analyze data for regression and classification. It represents different classes in a single plane iteratively to minimize the error and is also memory efficient.

It is one of the best algorithms for classification and regression applications. But in case of noisy datasets and large datasets, its performance dips because of chances of overlapping of classes.



Advantages

- 1) SVM calculation has a regularization parameter, which stays away from over-fitting.
- 2) SVM algorithm is memory efficient works effectively in high dimensional spaces.

Accuracy For SVM

Accuracy is the number of correctly predicted data points out of all data points.

$$Accuracy = \frac{True_{positive} + True_{negative}}{True_{positive} + True_{negative} + False_{positive} + False_{negative}}$$

Support Vector model accuracy (in %): 98.4090909090909

4.2 Random Forest Classifier

Random forest algorithm is a supervised machine learning algorithm. It will solve classification problems. It contains a number of decision trees for different subsets of the data instead of working on the whole data as a single subset. This improves the accuracy of prediction of the model by several folds as it takes the average of predictions of all the trees and decides the final output on the basis of majority votes of the predictions. This makes it suitable even for the large and varied datasets as it can deliver results with high accuracy in very less amount of time.

Advantages

- 1) The random forest algorithm will split the data into multiple trees and each tree will be trained on subset of data.
- 2) Random Forest algorithm is stable if a new data point is introduced in the dataset the overall algorithm is not affected.

Accuracy For Random Forest

Random Forest model accuracy (in %): 99.0909090909091

4.3 Naive Bayes Classifier

Naive Bayes classifier is used for classification problems and it is a probabilistic algorithm . It works based on Bayes theorem. It assumes that a particular feature is independent of the value of any other feature, given the class variable. The learners are used to predict the class labels. For uncertain data Hierarchical Naïve Bayes Classifier is used. To work with Naïve Bayes we have to split the dataset into two parts Feature Matrix and Response Matrix.

Advantages

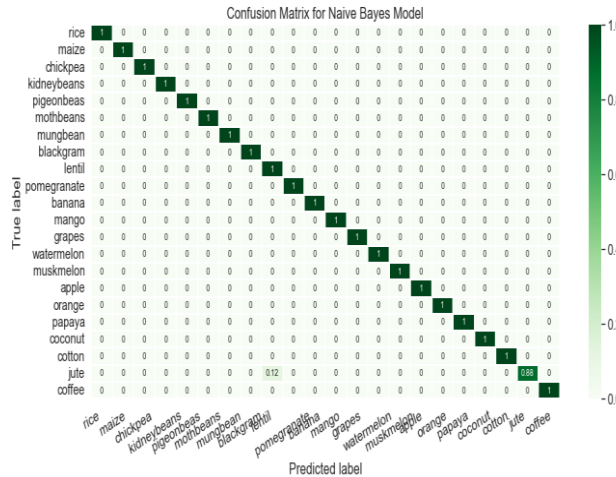
- 1) It is simple and easy to implement.
- 2) It doesn't require as much training data.
- 3) It handles both continuous and discrete data.

Accuracy For Naïve Bayes

Naïve Bayes model accuracy (in %): 99.43181818181817



Confusion Matrix For Naive Bayes



4.4 Scikit-Learn

Scikit-learn is an ML library for python. It has various algorithms for classification and regression like logistic regression, random forest classifiers and support vector machines. We can operate it along with other python libraries like NumPy and pandas. Scikit-learn is one of the best libraries especially for supervised learning which involves training the model by loading a sample dataset which it can observe and structure its learning accordingly. It also gives us the provision to use train_test_split for making training and testing datasets.

4.5 train_test_split Library

We import train_test_split from scikit-learn to split our dataset into two different sets according to our needs- one for training the model and the other for testing the working and accuracy of the trained model so that we can choose the best possible algorithm. We apply different supervised learning algorithms on the training data and obtain results on testing data for all the applied algorithms. Then the best performing algorithm is selected for the model.

	N	P	K	temperature	humidity	ph	rainfall
count	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000
mean	50.551818	53.362727	48.149091	25.616244	71.481779	6.469480	103.463655
std	36.917334	32.985883	50.647931	5.063749	22.263812	0.773938	54.958389
min	0.000000	5.000000	5.000000	8.825675	14.258040	3.504752	20.211267
25%	21.000000	28.000000	20.000000	22.789375	60.281953	5.971693	64.551686
50%	37.000000	51.000000	32.000000	25.598693	80.473146	6.425045	94.867624
75%	84.250000	68.000000	49.000000	28.561654	89.948771	6.923643	124.267508
max	140.000000	145.000000	205.000000	43.675493	99.981876	9.935091	298.560117

Figure 4. Explorative Analysis On Dataset

4.6 Working Algorithm Applied

The Jupyter Notebook is a tool used to develop data science projects. We started off with importing the required libraries that were pandas, matplotlib, seaborn and sklearn, which we felt were essential for carrying out proper analysis of the givendataset. The data is stored in a csv file. We then plotted a heatmap to check the correlation between all the factors.



Data splitting contains two parts: split part and train part. We then trained the model with linear regression and verified the accuracy with the testing data. The same was done for SVM and Random Forest Regressor also. The model with the best results, i.e., the Naïve Bayes Classifier with 99.4% accuracy was selected to make the prediction model. The application predicts the crops accurately.

V. RESULTS AND DISCUSSION

We have used a sample data set from Kaggle, which consisted of various factors needed for the proper growing of a crop.

The screenshot shows a web form for fertilizer recommendation. It contains three input fields: 'Nitrogen' with the value 90, 'Phosphorous' with the value 42, and 'Pottasium' with the value 43. Below these is a dropdown menu labeled 'Crop you want to grow' with 'maize' selected. A blue button labeled 'Predict' is positioned at the bottom center of the form.

Figure 5. Fertilizer Recommendation

The output will be like this

The K value of your soil is high.
Please consider the following suggestions:

1. Loosen the soil deeply with a shovel, and water thoroughly to dissolve water-soluble potassium. Allow the soil to fully dry, and repeat digging and watering the soil two or three more times.
2. Sift through the soil, and remove as many rocks as possible, using a soil sifter. Minerals occurring in rocks such as mica and feldspar slowly release potassium into the soil slowly through weathering.
3. Stop applying potassium-rich commercial fertilizer. Apply only commercial fertilizer that has a '0' in the final number field. Commercial fertilizers use a three number system for measuring levels of nitrogen, phosphorous and potassium. The last number stands for potassium. Another option is to stop using commercial fertilizers all together and to begin using only organic matter to enrich the soil.
4. Mix crushed eggshells, crushed seashells, wood ash or soft rock phosphate to the soil to add calcium. Mix in up to 10 percent of organic compost to help amend and balance the soil.
5. Use NPK fertilizers with low K levels and organic fertilizers since they have low NPK values.
6. Grow a cover crop of legumes that will fix nitrogen in the soil. This practice will meet the soil's needs for nitrogen without increasing phosphorus or potassium.

We have taken into consideration these factors which will be taken as an input from the farmer/authority that will be deciding which crop to sow. A detailed analysis has been carried out on these factors and various inferences were generated from the results that we then took into consideration while predicting the optimum crop.



```
1 model.predict([[74,35,40,26.491096,80.158363,6.980401,242.864034]])
array(['rice'], dtype=object)

['rice' 'maize' 'chickpea' 'kidneybeans' 'pigeonpeas' 'mothbeans'
 'mungbean' 'blackgram' 'lentil' 'pomegranate' 'banana' 'mango' 'grapes'
 'watermelon' 'muskmelon' 'apple' 'orange' 'papaya' 'coconut' 'cotton'
 'jute' 'coffee']
```

Figure 6. Unique Crops Labeled by Algorithm

The dataset has been analyzed and compared by using different diagrams like bar charts, histogram and pie chart etc. The attributes taken into consideration are the ratio of nitrogen in soil, ratio of phosphorous in soil, ratio of potassium in soil, temperature, humidity, pH of soil and the amount of rainfall. This covers varied factors which affect the growth of crops and are the deciding factors in which crop to be chosen to sow.

The model then shows the label as output which is the recommended crop. After applying three different algorithms- logistic regression, SVM and Random Forest Classifier, we found out that Random Forest Classifier gave the best results and so the model has been made using the same.

Nitrogen

Phosphorous

Pottasium

ph level

Rainfall (in mm)

State

City

Figure 7. Crop Recommendation

The Output will be like this

You should grow *coffee* in your farm

VI. FUTURE ENHANCEMENT

The project is very useful for farmers in present and future. The technologies will be upgraded and we can change the data as per our requirements. As we know that we are moving towards technological world, agricultural sector also moving towards new technologies. This project is developed as a web application and we can access this through phones, laptops and desktop. We can also use IOT tools to monitor crops. This will help the future farmers to grow crops which will give better yield.



VII. CONCLUSION

India's economic growth mainly focuses on agriculture. By using Machine Learning Algorithms we are predicting the crop which will help the farmers to get more profit. Thus the farmer's can plant the right crop which increases their yield and also increases the overall productivity of the nation. In future we want to improve the data set and also predict the crop yield.

REFERENCES

- [1] Rushika G., Juilee K, Pooja M, Sachee N, and Priya R.L.(2018). Prediction of Crop Yield using Machine Learning, Issue 02 IRJET (pg 2337-2339).
- [2] Ruchita T, Shreya B, Prasanna D, and Anagha C (2017). Crop Yield Prediction using Big Data Analytics, Volume 6, Issue 11, IJCMS.
- [3] Monali P, Santosh K, Vishwakarma, and Ashok V (2015). Analysis of Soil Behaviour and Prediction of Crop Yield using Data Mining Approach, ICCICN 2015.
- [4] Mrs. N. Hemageetha, and Dr. G.M. Nasira (2016). Analysis of Soil Condition based on pH value using Classification Techniques, IOSR-JCE, Issue 6, (pp.50-54).
- [5] E. Manjula, and S. Djodiltachoumy (2017). Data Mining Technique to Analyze Soil Nutrients based on Hybrid Classification. IJARCS.
- [6] Rohit Kumar Rajak, Ankit Pawar, MitaleePendke, Pooja Shinde, Suresh Rathod, and AvinashDevare (2017). Crop Recommendation System to maximize Crop yield using Machine Learning. Issue 12 IRJET.
- [7] S. Veenadhari, Dr. Bharat Misra, Dr. CD singh (2014). Machine Learning Approach for Forecasting Crop Yield based on Climatic Parameters. ICCCI-2014



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details