International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 04 | Apr 2022 www.irjet.net p-ISSN: 2395-0072

Crop Recommendation System Using Machine Learning Algorithms

Atharva Jadhav[1], Nihar Riswadkar[2], Pranay Jadhav[3] ,Yash Gogawale[4]

[1 2 3 4] *Student, Department of Computer Engineering, PES Modern College of Engineering, Pune, Maharashtra, India*

Abstract – *In today's time technology is playing a vital role in different sectors to overcome the difficulties and to have better and maximum results. In India, the farming sector has a huge impact on Indian economy. Half of the country's population is still employed in the agriculture sector. Agriculture industry is largely influenced by natural conditions of its surroundings and hence faces number of challenges in actual farming practices. Agriculture practices in the country are largely primitive and technological change in the sector is slow. Effective technology can be used to increase the yield and to reduce the maximum possible challenges in this field. Most of the times it is observed that farmers tend to sow the crop according to its market value and possible financial profits rather than taking factors like soil conditions, sustainability etc. in to the account. This may lead to undesirable results for farmers and for the nature of soil too. In today's time, technologies like machine learning and deep learning can become game changers in such fields if they are used in a proper manner. This paper will represent an effective use of such technologies in order to provide maximum assistance to farmers in the area of crop recommendation.*

Key Words: **Crop Recommendation, Machine Learning, Random Forest, Decision Tree, Logistic Regression, XGBoost, Data Analysis, Data Visualization**

1. INTRODUCTION

In India farming is not considered as a business but also has a huge impact on the social life of people which are associated to it. There are many festivals and social gatherings are celebrated in accordance with the different seasons and practices involved in farming. Hence large part of population is dependent on agriculture field directly or indirectly. But the situation of farmers is not good in India. Despite employing half the population in agriculture, the agriculture sector contributes to only 20% of Indian GDP. Hence, it is in dire need of improvement in order to make a good and profitable yield and also, a practical need without harming the nature. That's where technology comes in and can have major effects on agricultural sector. Basically, our project aims to tackle the difficulties faced by the farmers and aims to provide a correct crop for the farmers to grow and avoid the undesirable results by providing effective solutions using machine learning techniques.

2. Literature Review

The following research papers were referred by us before doing our project. While referring each of these papers we have come across various different findings discussed below. Each of the below entries discusses the title of the paper, the algorithms used and a general conclusion drawn from that research paper.

Title: Crop Suitability and Fertilizers Recommendation Using Data Mining Techniques
Author: Archana Chougule, Vijay Kumar Jha and Debajyoti Mukhopadhyay
Method Used: Random Forest, K- means clustering algorithm
Remark: Accuracy of random forest is found to be higher than ID3(Iterative Dichotomiser 3) algorithm for crop prediction and K- means for fertilizer recommendation.

Title: Random Forest Algorithm for Soil Fertility Prediction and Grading Using Machine Learning
Author: Keerthan Kumar T G, Shubha C, Sushma S A
Method Used: Random Forest, Gaussian Naïve Bayes, Support Vector Machine, Linear Regression
Remark: In case of Crop Prediction, Random Forest proves to be a better classifier as compared to Gaussian Naïve Bayes and Support Vector Machine while linear regression works efficiently for grading soil.

Title: Soil Classification using Machine Learning Methods and Crop Suggestion Based on soil series
Author: Sk Al Zaminur Rahman, S.M. Mohidul Islam, Kaushik Chandra Mitra
Method Used: Weighted K-NN, SVM, Bagged Tree
Remark: SVM has given the highest accuracy in soil classification as compared to K-NN and Bagged tree algorithms.

Title: AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms
Author: Zeel Doshi, Subhash Nadkarni, Rashi Agrawal, Prof. Neepa Shah
Method Used: Decision Tree, Random Forest, K- NN
Remark: Accuracy rates were Decision Tree (90.20), K-NN (89.78), Random Forest (90.43)

3. Methodology

While implementing the project, the following steps were implemented in order to achieve the results.

Data Cleaning and Preprocessing

One of the first steps is to make sure that the dataset we are using is accurate. The dataset should not have any missing values and if the dataset does have missing values, they

should be replaced by the appropriate values. The data should also be checked to see if there is a normal distribution for its features. The outliers should be removed. The skew value of the features should be checked and if the features have skewness, then those features should be normalized by using transformations. The dataset which we used had features having skewness in them. To normalize them, we have used quantile transformation on the features of our dataset.

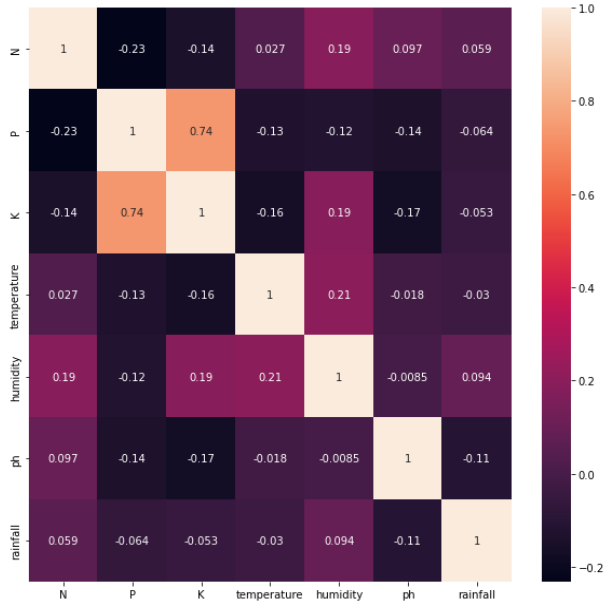
Data Analysis and Visualization

After performing the cleaning and preprocessing of the data, we perform data analysis and visualizations on our dataset. We try to analyze our data more clearly to find any trends or patterns in the dataset. We have created several visualizations of our dataset in order to understand the data properly. We have created bar charts, scatter plots, box plots etc. in order to visualize the data and find if there are any trends or patterns which we can find that will be useful while implementing our project.

Feature Selection

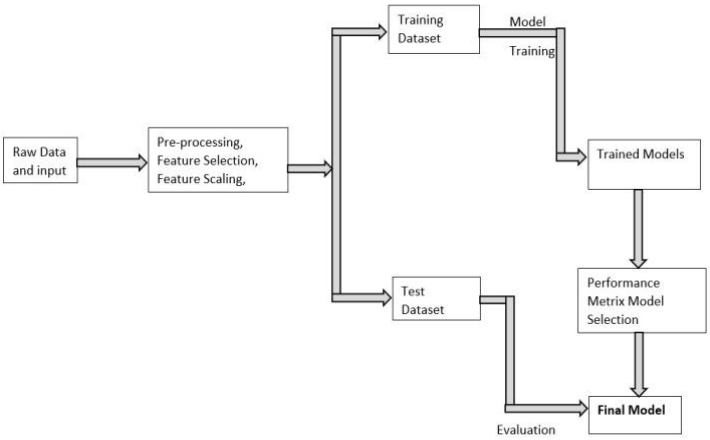
It is important that we select only those features that will be necessary to determine the type of crop to grow. For this, we have created a correlation matrix that shows the linear relationship of a feature with every other features. If features are highly correlated then that feature should be dropped, but as we can see in the below matrix that the features are not highly correlated with each other, hence it makes sense not to drop any of them and hence we will be using all of them to predict the type of crop to grow.

Here is the correlation matrix for our features in our dataset.



Model Building

The next step is building the machine learning model. While building the machine learning model, first we need to split our dataset into 2 parts i.e.: training data and test data. We have split the data in the ratio of 70-30. Taking the training data, we apply our machine learning algorithms on the features of the dataset. We have used 4 machine learning algorithms on our training dataset and the algorithms that gives us the highest accuracy will be selected on the test dataset.

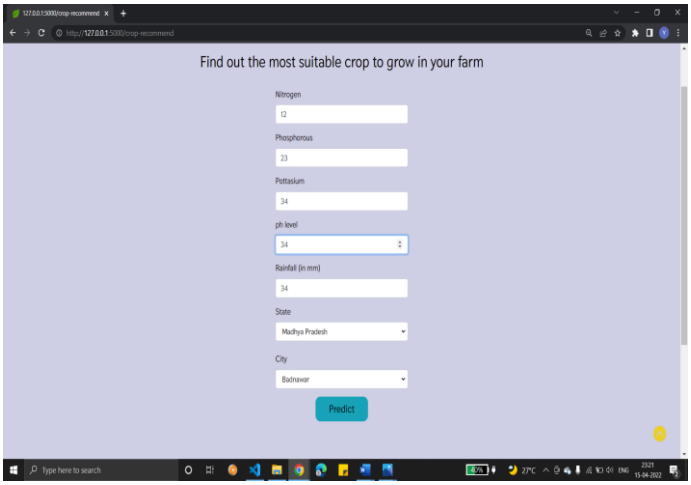


Architecture diagram of our machine learning model

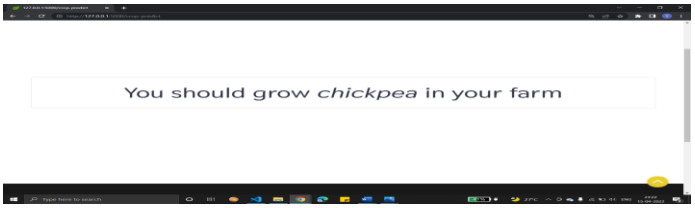
Building a UI

In the next step, we have built a UI for a user to input his data so that once he enters the information such as N, P, K values of soil, temperature, humidity, rainfall etc., the model will process the data and will recommend the appropriate type of crop to be grown in such a condition. Once the user enters the following values and submits the machine learning model will predict the crop that the person must grow.

Below is a screenshot of the UI which we have made for our project.



After putting the values by the user, we get the results of the model in the way mentioned below.



4. Dataset

The dataset for this topic was taken from Kaggle. It is a crop recommendation dataset giving us information about various types of crops and the features that decide which crop is suitable for growing.

Dataset Link:

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

Features of the Dataset

N: ratio of Nitrogen content in soil
P: ratio of Phosphorous content in soil
K: ratio of Potassium content in soil
Temperature: temperature in degree Celsius
Humidity: relative humidity in %
pH: pH value of the soil
rainfall: rainfall in mm

5. Machine Learning Algorithms Used

Random Forest

Random Forest is a supervised ensemble machine learning algorithm used in both classification as well as regression problems. It contains various decision trees and an average of it is taken so as to give the output. It is based on the concept of bagging wherein multiple decision trees are created and an average of them is taken so as to give the output. As decision tree are prone to overfitting, random forest is useful in reducing the effect of overfitting and hence giving a more accurate output.

Decision Tree

Decision Tree is one of the most popular machine learning algorithms used mostly in classification problems but can also be used for regression type of problems. The working of it is based on a simple technique, wherein a yes/no question is asked and according to the answer the tree is split in smaller nodes. The split of the nodes can either happen by calculating Gini impurity (calculates the measure of impurity) or information gain (calculates the change in the entropy). Decision Trees are prone to overfit and hence this may lead to getting a lower accuracy. This problem can be solved by using random forest algorithm.

Logistic Regression

It is one of the simplest algorithms in machine learning. It is used for solving classification problems. It uses a sigmoid function to mathematically calculate the probability of an observation and accordingly, the observation is then put into its respective class. While calculating, if the probability of an observation is 0 or 1, a threshold value is decided upon and classes having probabilities above the threshold value are given the value 1 and classes having values below the threshold are given the value 0.

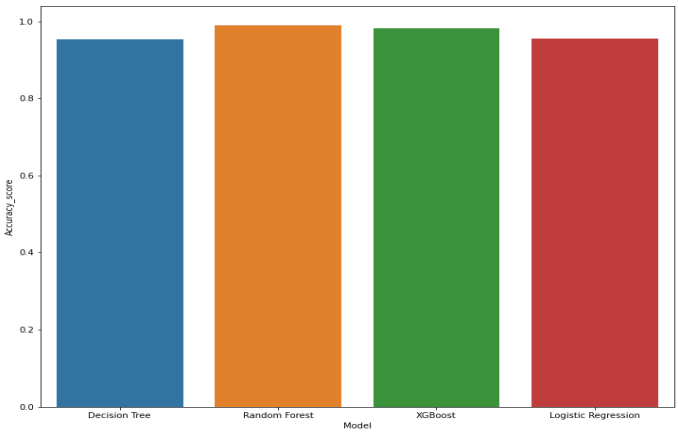
XGBoost

XGBoost is one of the most popular algorithms used today. It is a tree-based algorithm using gradient boosting framework. This algorithm is based on a feedback approach, where feedback from the decision tree is taken so as to improve the decision tree further (also known as boost) as it "boosts" the efficiency of the tree and helps in bringing out a better accuracy.

6. Results

After performing the data cleaning and visualizations, we implemented our machine learning algorithms on the features of the dataset. The four algorithms we have used are Decision Tree, Random Forest, Logistic Regression and XGBoost. The features selected of the dataset are N (Nitrogen), P (Phosphorous), K (Potassium) values of the soil, temperature, humidity, rainfall and pH value.

After implementing the four algorithms on our dataset, we can see that Random Forest gives us the highest accuracy out of all the four algorithms (98.9 %), followed by XGBoost (98.2 %). Logistic Regression gave an accuracy of 95.6% while Decision Tree gave the least accuracy at 95.3%.



Graphical Comparison of accuracies of all the Machine Learning Models

Model	Accuracy score
Random Forest	0.989394
XGBoost	0.981818
Logistic Regression	0.956061
Decision Tree	0.953030

7. Future Scope

At present we have limited our scope only to a crop recommendation system, but this in the future can be extended to other areas such as fertilizer recommendations, wherein a person can get the information of the fertilizer suitable for his crop. Another extended application of this could be plant disease classification using CNN and also giving a remedy to that disease. As agriculture is quite an unexplored area, the scope for such a type of project is tremendous. While we have made a working website for our model, in the future this can also be made into an app so that it is easier for the farmers to access. Both the website and the app can be made in regional languages so that people would be more comfortable using them.

8. CONCLUSION

As our farmers presently are not using technology in their farming practices, they are lagging behind in their farming approaches. Hence, we have made this project so as to encourage farmers to use the current technology instead of relying on old methods. Growing a crop requires a lot of knowledge and insight into many things such as the contents of soil, temperature of the place, pH of soil etc. and hence it would make sense for the farmer to adopt to newer technologies as it would make his life easier. Using machine learning algorithms, we predicted an accuracy of 98% of predicting the right crop to grow and if the farmer adopts to this technology which would not only make his life easier but also help him in making decisions that do not exploit the environment.

REFERENCES

[1] Keerthan Kumar T G, Shubha C, Sushma S A" Random Forest Algorithm for Soil Fertility Prediction and Grading using Machine Learning "International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-1, November 2019

[2] Archana Chougule, Vijay Kumar Jha and Debajyoti Mukhopadhyay "Crop Suitability and Fertilizers Recommendation using Data Mining Techniques" Springer Nature Singapore Pte Ltd. 2019

[3] Sk Al Zaminur Rahman, S.M. Mohidul Islam, Kaushik Chandra Mitra" Soil Classification using Machine Learning Methods and Crop Suggestion Based on Soil Series" 2018 21st International Conference of Computer and Information Technology (ICCIT), 21-23 December, 2018

[4] Zeel Doshi, Subhash Nadkarni, Rashi Agrawal, Prof. Neepa Shah "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms " 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)

[5] Mrs. N. Saranya, Ms. A. Mythili" Classification of Soil and Crop Suggestion using Machine Learning Techniques" International Journal of Engineering Research & Technology (IJERT) http://www.ijert.org IJERTV9IS020315 (4.0 International License.) Vol. 9 Issue 02, February-2020