Predicting Agricultural Produce using Machine Learning Techniques

Riddhi Dange[1], Jinisha Kande[2],Kunal Bhosale[3], Prof. Ankush Hutke[4]
[1,2,3] *Student, Department of Information Technology, MCT Rajiv Gandhi Institute of Technology*
[4]*Faculty, Department of Information Technology, MCT Rajiv Gandhi Institute of Technology*

***Abstract*—Agriculture is the largest contributor of the GDP in our country. But still the farmers do not get theexact worth price of the crops. It is mostly due to improper irrigation or imprecise crops selection, sometimes the crop yield is less than expected. In order to add efficiency to the whole agriculture process a "Predicting Agricultural produce using Machine Learning Techniques" is proposed in the following project. The system aims at predicting the agricultural produce by obtaining area on Google Maps API and further calculate area under cultivation.GPS technology is now used widely by various people and organizations for determining the position of objects. GPS can be used to measure the area of land under cultivation.**

***Index Terms*—Agriculture, Google Maps API, GPS, Machine Learning.**

I. INTRODUCTION

India is an agriculture-based developing country and India's economy predominantly depends on agriculture yield grow than d agro-industry products. India possesses a potential to be a superpower in the field of agriculture as the agricultural sector accounts for 18 percent of India's gross domestic product (GDP) and provides employment to 50% of the country's universal system to assist farmers in agriculture workforce. Thus, it is very essential to improve the productivity and efficiency of agriculture by providing improved methods of cultivation to the farmer. However, as rapidly the conditions change day by day, farmers are forced to cultivate more and more crops to keep up with the market. Following this as the current situation and their limited knowledge, many of them do not know about the new crops being cultivated, or crop recommended for the field conditions and are also not completely aware of the benefits they get while farming them. Thus, to improve efficiency of agriculture for farmers, the following project proposes a system in which the first part, crop prediction module predicts agriculture produce by obtaining the coordinates of the crop field on Google Maps API using GPS which is then uploaded in the system and used to calculate the area under cultivation using marked points on maps. Further, the crop name and area under cultivation will be used to predict the crop yield and this information is provided to farmers. A well-performing model for crop recommendation is essential in the domain of Agriculture to assist Farmers. An ideal system is accurate only when it's trained on an easily available, reliable dataset, shows real-time predictions, and correctly classifies various types of crops. For our research purposes, we have implemented a "Prediction of Agriculture produce using Machine Learning Techniques" project which utilizes Random Forest Classifier to recommend the most profitable crop for the following parameters: Nitrogen, Phosphorus, Potassium content, Temperature, Humidity, pH value and Rainfall. The input for these parameters is taken from the user via a user interface in form of a website and the recommended crop is given as the output.

II. PROPOSED SYSTEM

Human error and misjudgment can lead to a wastage of crop produce due a misconceived need for a certain type of crop over another by farmers as well as the incorrect pricing of crops by administrative officers. Hence, we propose an idea to identify the sustainability of crops in a particular land area based on certain parameters which may improve the overall efficiency of the agriculture sector. Our project primarily focuses on two modules. One module depends on the Prediction of Agricultural Produce. In the second module we will look into Real-Time Crop Recommendation.

A. Crop Prediction

In this module we start by using Google Maps API to calculate the area under cultivation using mapped points on maps. Then, we use the widely available datasets from the Ministry of Agriculture and Farmers Welfare, India and identified crop name and area under cultivation will be used to predict the crop yield. Google Maps API is used to calculate the area under cultivation using mapped points on maps. This part of the module utilizes a modified version of Shoelace Algorithm.

B. Crop Recommendation

This is basically a prototype for a real-time crop recommendation algorithm using Machine Learning and Data Analytics. The proposed system will integrate the data obtained from soil, crop repository, weather department and by applying machine learning algorithm: Random Forest Regression, a prediction of most suitable crops according to current environmental conditions is made.

Thus, the proposed system will intelligently help farmers get maximum crop output and knowledge. Hence, our prime focus of providing farmers a tool which will improve the agricultural output is also fulfilled.

III. METHODOLOGY

A. Google Maps API

Google maps API Google Maps APIs are prepackaged pieces of code which let users easily include maps on platforms such as websites or mobile apps, and also add functions to applications. In the project, we use Google Maps API for area calculation module to calculate the area for cultivation by using mapped points.

The Maps JavaScript API lets a user manually customize maps with content and imagery which can be displayed on technologies suc as web pages and mobile devices. This technology features four basic map types: roadmap, satellite, hybrid, and terrain, which can be modified using layers and styles, controls and events, and various services and libraries.

B. Random Forest Algorithm-

Random Forest Classifier-

Random forest, consists of many individual decision trees which operate a as an ensemble wherein each individual tree in the random forest spits out a class prediction and the output with the most votes become our model's prediction.

Random Forest Regression-

Random Forest Regression being a supervised learning algorithm which uses ensemble learning method for regression. Ensemble learning method is a technique that combines predictions from algorithms, in this case, predictions of a decision trees, to make a more accurate prediction than a single model and giving output as the average of all decision tree results.

IV. IMPLEMENTATION

A. Area Calculation

The area calculation feature is used to calculate the area of any location using the Google Maps API. Mark points on the map of the area you want to calculate, once the points marked are a closed polygon i.e., a closed figure then the area will be calculated and displayed. The area is calculated in meter square. Below is a representation of how this module looks like.

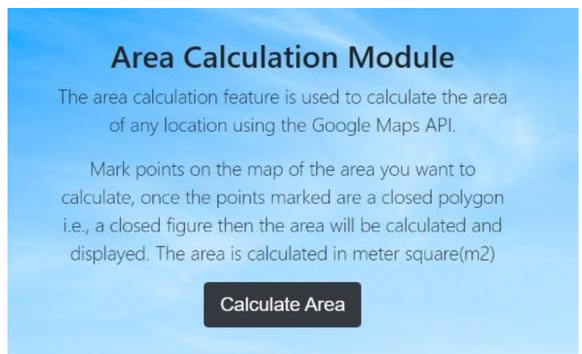

Fig.1 Area Calculation Module

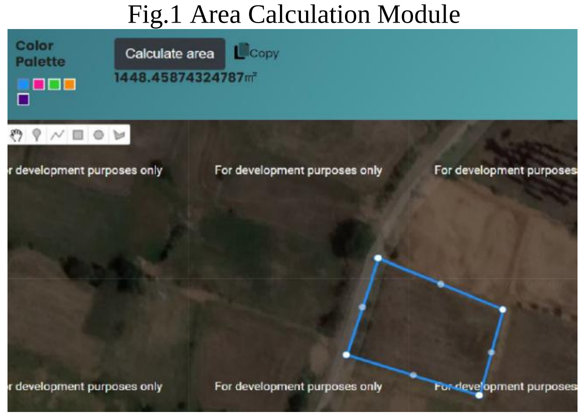

Fig.2 Area Calculation Demonstration

A drawing manager is attached on the map to let the user draw markers, lines and shapes. Once the marked points or lines are a polygon then the calculated area is displayed on the screen. To calculate the area, the Google Maps Geometry Library is used. The function computeArea() is responsible for the calculation of the area. The function is written in the “area_map.js” file.

B. Crop Recommendation

In this system, the user can use the soil data from their side and based on factors like NPK values, temperature, humidity, pH and rainfall. Using the Random Forest Classifier the prediction of the suitable crop for the suitable conditions will be achieved. This dataset was obtained from Ministry of Agriculture and Farmers Welfare.The dataset will be explained thoroughly as we go further.

	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

2200 rows × 8 columns

Fig.3 Dataset

Parameters-

N, P, K-

All crops need nitrogen, phosphorus, and potassium for their growth.

Nitrogen (N) – Nitrogen is accountable for the growth of leaves on the plant.

Phosphorus (P) – Phosphorus helps in root growth, flower andfruit development.

Potassium (K) – Potassium helps the overall functions of the plant to function.

Temperature-

Seed germination, root function, rate of plant growth, as well as occurrence and severity of plant diseases are affected by soil temperature.

Humidity-

Presence of high humidity may promote the growth of mold and bacteria which causes plants and crop yield to fail, as well as conditions like root or crown rot.

pH-

A soil pH below 5.6 is generally considered to be low for most crops production. An ideal pH range is considered to be in between 6.0 and 7.0

Rainfall-

In the south-west portion of India, monsoon irrigates over half of India’s crop land. This arrival signifies the beginning of the cultivation of rain-fed kharif crops majorly dependent on the monsoon. Therefore, this parameter plays an important role for determining a crop to be produced.

Separating features and target variables-

Features are the individual independent variables which act as the input in the system. For our dataset, the NPK values (all individually i.e N, P, K), Temperature, Humidity, pH and Rainfall are the “features”.

The target is the output of the input variables. This could be the individual classes that the input variables maybe mapped to in case of a classification problem. In our model, the target is the Crop to be produced.

```
features = df[['N', 'P','K','temperature', 'humidity', 'ph', 'rainfall']]
target = df['label']
#features = df[['temperature', 'humidity', 'ph', 'rainfall']]
labels = df['label']
```

Fig.4 Target and Feature Label

Splitting into train and test data-

Splitting dataset into training an testing. The idea of a dataset to be “sufficiently large” is specific to each predictive modeling problem. This means that there is enough data to be split into train and test datasets and each of these train and test datasets are suitable representations of the problem domain. This requires that the original dataset is also a suitable representation of the specific problem domain.

```
# Splitting into train and test data


from sklearn.model_selection import train_test_split
Xtrain, Xtest, Ytrain, Ytest = train_test_split(features,target,test_size = 0.2,random_state =2)
```

Fig.5 Splitting into train and test data

Using Random Forest Classifier-

The sklearn. ensemble module includes two algorithms out of which one is Random Forest Classifier. This algorithm is specially designed for trees.

The train and test data that we acquired will be used here. The Accuracy Score using the “Random Forest Classifier” for the train and test data is 0.99090.

```
from sklearn.ensemble import RandomForestClassifier

RF = RandomForestClassifier(n_estimators=20, random_state=0)
RF.fit(Xtrain,Ytrain)

predicted_values = RF.predict(Xtest)

x = metrics.accuracy_score(Ytest, predicted_values)
acc.append(x)
model.append('RF')
print("RF's Accuracy is: ", x)
```

```
RF's Accuracy is:  0.990909090909091
```

Fig.5 Random Forest Classifier

Making a prediction-

Creating an array "data", which will consist of all the features we need for the target prediction.

The values in the list is [N, P, K, Temperature, Humidity, pH, Rainfall].We imported Numpy as np.

Np.array(data) converts list into an array so now we have the array "data". The array is then used in predicting the crops under the input conditions given.

The "RF.predict(data)" is used to predict the recommended crop. RF is the Random-Forest-Classifier algorithm. As shown in the below figure the crop recommended for given conditions is Jute.

```
data = np.array([[83, 45, 60, 28, 70.3, 7.0, 150.9]])
prediction = RF.predict(data)
print(prediction)
```

```
['jute']
```

Fig.6 Random Forest Algorithm

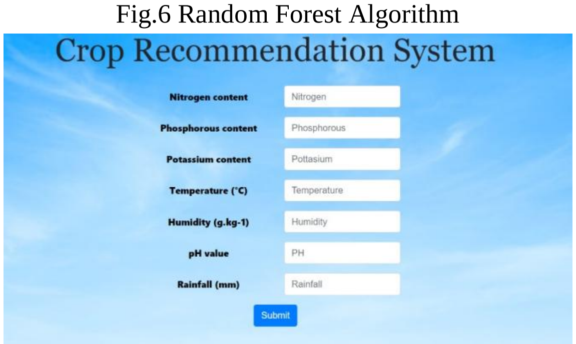

Fig.7 Crop Recommendation System

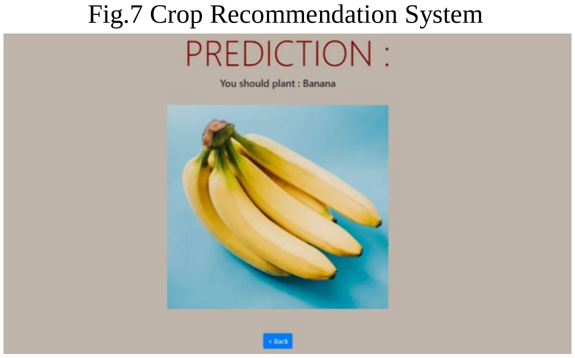

Fig.8 Crop Recommendation Output

C. Crop Prediction

Connecting our first module (Area Calculation Module) to the Crop Prediction Module, we will be using the output of the first module as an input in this module.

In this system, the user can use the current district, the season he prefers and the area of your field.

Understanding the inputs:

Current District-

The current district could be the district of the respective farmer/user where the district strictly belongs to the state of Maharashtra. The input of district depends totally upon the user input.

Current Season-

The seasons taken under consideration for the prediction are mainly, Kharif and Rabi.

The season input depends upon for which season the farmer wants to predict its probable yield for the crops of the selected season.

Area of field-

Area of field is as mentioned before, the area which can be calculated using the "Area Calculation Module".

The area calculated is in "m^2". The user is free to add his/her already calculated area but it should be in the unit mentioned before. Prediction according to the area, district and season preferred.

Fig. 9 Crop Prediction Module

The output gives you the crops which can be produced in a particular season, in a particular district as well as the area of the field of the farmer.

The output is in a decreasing manner i.e. the highest production of a crop will be on top.

Prediction of crops

Serial Number	Crop Name	Production (in tonnes)
0	Wheat	2047.12
1	Sunflower	1186.0
2	Maize	1155.0
3	Gram	1154.0
4	Rapeseed &Mustard	1154.0
5	Soyabean	1154.0
6	Safflower	1151.0
7	Jowar	1134.0

Fig. 10 Crop Prediction Output

For example
As seen in the above figure, the inputs are District is Akola, Season is Rabi and Area of field is approximately 1000 m^2.
The output crops include, Wheat, Sunflower, Maize, etc with Wheat being the highest production i.e. 2047 tonnes and Jowar being the lowest i.e. 1134 tonnes.

V. RESULTS

A. Summary-
In summarization of our work, we've implemented in two parts, i.e. creating an area calculator for land area calculation in meters and a crop recommendation system in the form of an interactive website. The interactive website we've made would also be very helpful in land area detection, fast crop recommendation and prediction.

B. Conclusion-
In this project, we aim to use Random Forest Algorithm to provide us with accurate crop recommendations when the user puts in the inputs required. By the means of Google maps API we are also getting a precise measure of the land in concern. By this project we aim to achieve test results having high accuracy in recommendation and prediction of crops as well as provide the accurate area of land under cultivation. Thus, it makes it helpful to the farmers in both sectors.
This undertaking attempts to start an area that is in great demand. During this project, we investigated different papers and we picked up information on machine learning algorithms and we got various ideas from these papers. Random Forest Regression algorithm as well as Google maps API for area selection was an inspiration from all the papers we went through. We aspire for the outcomes and strategies introduced in this article to be further extended to a greater task.

C. Recommendation for Future study-
The future extent for this work includes a hold in farm planning. The work proposed is also quite useful for the detection of appropriate crop and fertilizer according to the crop. This project can assist farmers in calculating the precise area of their land. The crop recommendation module can be expanded to include more variety of crops.

D. Concluding Remarks-
We found Random Forest Algorithm gave a very quick output and it clearly outperforms traditional machine learning techniques owing to its prediction of outputs with high accuracy, even for large datasets.

REFERENCES

[1] E. Manjula and S. Djodiltachoumy, “A Model for Prediction of Crop Yield”, International Journal of Computational Intelligence and Informatics, Vol. 6: No. 4, March 2017
[2] Jérôme Treboux and Dominique Genoud, “Improved Machine Learning Methodology for High Precision Agriculture”, 2018 Global Internet of Things Summit (GIoTS).
[3] V. A. Windarni, E. Sediyono, and A. Setiawan, “Using GPS and Google maps for mapping digital land certificate,” Informatics and Computing (ICIC), International Conference on, 2016, pp. 422–426.
[4] Nischitha K , Dhanush Vishwakarma, Mahendra N, Ashwini, Manjuraju M.R, “Crop Recommendation using Machine Learning Approaches”, International Journal of Engineering Research & Technology (IJERT), www.ijert.org Vol. 9 Issue 08, August-2020.
[5] Shruti Mishra Priyanka Paygude Snehal Chaudhary Sonali Idate , “Use of Data Mining in Crop yield prediction, Proceedings of the Second International Conference on Inventive Systems and Control (ICISC 2018) IEEE Xplore.
[6] Deepak Murugan, Akanksha Garg, Tasneem Ahmed, and Dharmendra Singh, “Fusion of Drone and Satellite Data for Precision Agriculture Monitoring”, 2016, 11th International Conference on Industrial and Information Systems (ICIIS).
[7] Dmitrii Shadrin, Andrey Somov, Tatiana Podladchikova, and Rupert Gerzer, “Pervasive Agriculture: Measuring and Predicting Plant Growth using Statistics and 2D/3D Imaging”, 2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC).
[8] Prof. D.S. Zingade, Omkar Buchade, Nilesh Mehta, Shubham Ghodekar, Chandan Mehta, “Crop prediction system using machine learning”, International Journal of Advance Engineering and Research Development.

[9] Vikky Aprelia Windarni, Eko Sediyono, Adi Setiawan, “Using GPS and Google Maps for Mapping Digital Land Certificates”, 2016 International Conference on Informatics and Computing (ICIC).

[10] Rushikesh Bhav, Mayuresh Deodhar, Kevin Bhalodia, Prof. Mansing Rathod, “Android Application for Crop Yield Prediction and Crop Disease Detection”, International Journal of Innovative Science and Research Technology, Volume 3, Issue 3, March– 2018.

[11] Santosh Reddy,, Abhijeet Pawar, Sumit Rasane, Suraj Kadam , “ A Survey on Crop Disease Detection and Prevention using Android Application”, IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 2 Issue 4, April 2015.