International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 1 Jan. 2023, pp: 1104-1109 www.ijaem.net ISSN: 2395-5252

Smart Farming Using Machine Learning: A Focus on weather based Crop Selection

NayanaN.Ghuikar , Prof. Ashok P.Kankale
Department of Computer Science and Engineering, Rajshri Shahu College of Engineering Buldhana 443001, Maharashtra India

---

Date of Submission: 17-01-2023 Date of Acceptance: 31-01-2023

ABSTRACT-- The Indian economy relies heavily on agriculture. However, agriculture in India is currently undergoing structural change, resulting in a crisis. The only way to get out of this mess is to do everything in our power to turn farming into a profitable business and get farmers to keep growing crops. This research would use machine learning to assist farmers in making appropriate decisions regarding cultivations as an effort in this direction. Using supervised machine learning algorithms, this project focuses on predicting the appropriate crop based on climatic conditions and the crop's yield from historical data. Based on weather and soil parameters, this project will propose a crop selection strategy to maximize crop yield. Using seasonal weather forecasting, it also suggests the best time to sow suitable crops. Weather prediction is made possible by machine learning algorithms like the recurrent neural network, and crop selection is made possible by the Random forest classification algorithm.Keywords:Machine Learning, Crop, Decision Tree

I. INTRODUCTION

The Indian economy is based on agriculture. However, India's agricultural sector is experiencing a crisis as a result of structural changes. Over time, agriculture's relative contribution to GDP has steadily decreased. India's transition from food self-sufficiency to net food importation is troubling. The agricultural sector in India is currently facing a crisis, as indicated by all of these trends.

It is argued that the Indian agricultural crisis will have a significant impact on all other sectors and the nation's economy in multiple ways. The only way to get out of this mess is to do everything in our power to turn farming into a profitable business and get farmers to keep growing crops.

Farmers used to predict their yields based on the yields of previous years. As a result, we are able to predict crop yield using a variety of techniques or algorithms for this kind of data analytics in crop prediction. People in today's society are unaware of the importance of cultivating crops at the right time and location. The weather, temperature, and a number of other factors have all been examined, and there is no effective solution or technology to resolve the issue. When making decisions about agricultural risk management, accurate information about crop yield history is crucial. As a result, the idea presented in this paper is to predict crop yield and climatic conditions using historical crop data. Before cultivating an acre, the farmer will check the crop's production per acre. The primary goals are:

1. To make use of machine learning methods to predict crop yield and quality.
2. To properly process the data and analyze it in order to make more accurate predictions.
3. To make machine learning models work better.
4. To build a web application that is simple to use.

As a result, the idea presented in this paper is to predict the crop's yield and climatic conditions using historical crop data. Before cultivating the field, the farmer will check the crop's production per acre.]

II. PROBLEMSTATEMENT

Important agricultural issues include crop yield prediction and crop selection. Using machine learning algorithms, this project aims to predict a suitable crop based on the climate parameters and location, as well as the crop's yield based on the season and field area

III. LITERATUREREVIEW

In [1]. Machine learning approach for forecasting crop yield based on parameters of climate. The paper was provided at the International Conference on Computer Communication and Informatics (ICCCI). In the current research, a software tool named Crop

Advisor has been developed as a user-friendly web page for predicting the influence of climatic parameters on crop yields.C4.5 algorithm is used to produce the most influencing climatic parameter on the crop yields of selected crops in selected districts of Madhya Pradesh.
In [2].Analysis of Crop Yield Prediction by making Use of Data Mining Methods. IJRET: The paper provided in the
International Journal of Research in Engineering and Technology. In this paper, the main aim is to create a user-friendly interface for farmers, which gives the analysis of rice production based on the available data. For maximizing the crop productivity various Data mining techniques were used to predict the crop yield.

In [3]. Random Forests for Global and Regional Crop Yield Predictions. institute on the Environment, University ofMinnesota, St. Paul, MN 55108, United States of America. The generated outputs show that RF is an effective and different machine-learning method for crop yield predictions at regional and global scales for its high accuracy.

In[4] Crop Prediction using Machine Learning This research work helps the beginner farmer in such a way to guide them for sowing the reasonable crops by deploying machine learning. Naive Bayes, a supervised learning algorithm puts forth in the way to achieve it. The proposed supervised machine learning using naive Bayes Gaussian classifier with boosting algorithm is developed to predict the crop at high accuracy.
The Naïve bayes works on the basis of Bayes theorem.
P(C/X) = P(X/C) P(C) /P(X) ------(1)

- P(c|x) is the posterior probability of class (c, target) given predictor (x, attributes).
- P(c) is the prior probability of class.
- P(x|c) isthe likelihood which is the probability of predictor given class.
- P(x) is the prior probability of predictor. Based on the posterior probaility the future of data can be predicted.Naïvebayes work well for large data set.

In[5] Smart Farming Prediction Using Machine learning, the paper is about using machine learning with variousenvironmental factors like soil, pressure, weather, crop type to predict the maximized profitable crop to grow. The papermainly focuses on the algorithms used to predict crop yield, crop cost predictions.

In [6]Crop Prediction on the Region Belts of India: A Naïve Bayes MapReduce Precision Agricultural Model The planned work introduces an efficient degree economical crop recommendation system. From the yield graphs, the simplest time of sowing, plant growth, and gathering of plants may be known. Conjointly the best and worst condition may also be incurred. The model focuses on all styles of farms, and smaller farmers may also be benefitted. This model may be increased to seek out the yield of each crop, and for a chemical recommendation.

IV. ANALYSIS AND DESIGN

We made the decision to solely execute the system in India's Maharashtra State. To put the system into place, historical information regarding the district-level crop and climate was required. The information on this page, which covers State, District, Season, Crop, Area, and Production, was compiled from the government's website, www.data.gov.in.anddirectorate of economics and statistics, planning department, government of Maharashtra.Mumbai(https://bankofmaharashtra.in/writereaddata/documentlibrary) the climatic information
From Kaggle, data on temperature, humidity, soil pH, rainfall, and crop class label that are suitable for the specific crops have been gathered. The datasets used for this project are shown in the following snapshots.

```
= pd.read_csv('crop_prediction.csv')
head()
```

Temperature	Humidity	pH	Rainfall	Label
20.879744	82.002744	6.502985	202.935536	Rice
21.770462	80.319644	7.038096	226.655537	Rice
23.004459	82.320763	7.840207	263.964248	Rice
26.491096	80.158363	6.980401	242.864034	Rice
20.130175	81.604873	7.628473	262.717340	Rice

Figure. 1. Dataset For Crop Prediction Problem

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 1 Jan. 2023, pp: 1104-1109 www.ijaem.net ISSN: 2395-5252

```
sv('crop_yield_prediction.csv')
```

District_Name	Season	Crop	Area	Production
AHMEDNAGAR	Rabi	Maize	1	1113
AHMEDNAGAR	Kharif	Pigeon Peas	17600	6300
AHMEDNAGAR	Kharif	Chick Peas	40800	18600
AHMEDNAGAR	Kharif	Maize	4400	4700
AHMEDNAGAR	Kharif	Mung Beans	10200	900

Figure. 2. Dataset For Crop Yield Prediction Problem

Rainfall class (percentage to normal)	No. of talukas#									
	June		July		August		September		October	
	2020	2021	2020	2021	2020	2021	2020	2021	2020	2021
120 & above	177	231	113	177	144	35	180	298	220	150
100 – 120	64	58	51	64	58	22	46	13	24	46
80 – 100	50	32	69	57	60	48	58	13	41	54
60 – 80	34	17	58	39	31	107	42	19	35	44
40 – 60	26	12	42	15	39	91	25	9	26	31
20 – 40	4	5	21	3	3	46	4	3	7	23
0 – 20	0	0	1	0	0	8	0	0	2	7

Table1:Classification of talukas according to rainfall received

(mm)

Rainfall during			Region							
			Konkan#	Nashik	Pune	Aurangabad	Amravati	Nagpur	Maharashtra	
June		Normal	662.5	139.7	198.6	134.0	147.6	187.1	207.6	
	2020	Actual	541.2	210.0	201.0	209.2	183.1	193.3	235.1	(113.2)
	2021	Actual	932.4	126.6	234.6	200.9	212.7	237.7	282.7	(136.2)
July		Normal	1,063.8	238.5	327.2	186.2	259.6	362.3	330.9	
	2020	Actual	886.1	204.7	165.4	208.0	247.6	268.8	287.5	(86.9)
	2021	Actual	1,428.1	146.0	374.5	265.0	267.5	359.2	401.0	(121.2)
August		Normal	766.3	197.4	247.5	193.3	231.7	347.8	286.0	
	2020	Actual	1,222.0	192.6	310.4	151.1	199.5	444.4	353.6	(123.6)
	2021	Actual	305.8	136.3	86.1	174.2	163.6	192.8	173.6	(60.7)
September		Normal	375.5	151.5	171.7	166.0	130.9	175.3	179.7	
	2020	Actual	429.9	193.9	189.3	278.3	151.0	139.1	218.7	(121.7)
	2021	Actual	390.1	268.8	152.9	308.4	269.7	321.6	332.5	(185.0)
October		Normal	108.6	53.1	96.5	71.3	57.5	53.4	71.1	
	2020	Actual	242.5	71.0	232.7	118.7	98.2	51.5	124.7	(175.4)
	2021	Actual	102.1	74.9	89.2	61.9	87.6	34.0	81.0	(113.9)
June to October		Normal	2,876.7	760.2	1,041.5	758.8	807.3	1,125.9	1,075.3	
	2020	Actual	3,321.7	872.2	1,088.8	944.3	840.4	1,097.1	1,219.6	(113.4)
	2021	Actual	3,627.7	752.6	937.3	1,112.4	1,001.1	1,345.3	1,270.8	(118.2)

Table 2:Regionwise actual rainfall received (Note: Figures in bracket indicate percentage of actual rainfall to normal rainfall)

Crop	Area ('000 ha)			Production ('000 MT)		
	2020-21	2021-22 (tentative)	Per cent change#	2020-21	2021-22 (tentative)	Per cent change#
Rice	1,473	1,549	5	3,027	3,237	7
Jowar	379	209	(-)45	381	173	(-)55
Bajra	687	504	(-)27	906	458	(-)49
Ragi	82	73	(-)10	94	94	(-)0.4
Maize	810	873	8	2,625	2,350	(-)11
Other Cereals	37	44	18	17	17	(-)1
Total Cereals	3,467	3,253	(-)6	7,050	6,308	(-)11
Tur	1,340	1,335	(-)0.3	1,450	963	(-)34
Moong	401	377	(-)6	207	183	(-)12
Udid	356	436	22	227	235	4
Other Pulses	123	79	(-)36	90	68	(-)24
Total Pulses	2,220	2,226	0.3	1,974	1,448	(-)27
Total Foodgrains	5,688	5,480	(-)4	9,024	7,756	(-)14
Soyabean	4,290	4,617	8	6,264	5,422	(-)13
Groundnut	226	203	(-)10	271	240	(-)11
Sesamum	19	8	(-)61	5	2	(-)88
Nigerseed	8	6	(-)26	2	1	(-)30
Sunflower	19	12	(-)35	9	6	(-)31
Other Oilseeds	10	3	(-)72	4	1	(-)79
Total Oilseeds	4,573	4,849	6	6,554	5,672	(-)13
Cotton (Lint)$	4,545	3,954	(-)13	10,110	7,112	(-)30
Sugarcane**	1,143	1,232	8	1,11,642	1,11,200	(-)0.4
Total	15,948	15,515	(-)3	-	-	-

Table 3:Area and production of principal kharifcrops

Kharifcrops:

During kharifseason of2021-22, the sowing was completed on155.15 lakh ha, as against 159.48 lakh haduring previous year in the State. The areaunder pulses, oilseeds and sugarcane isexpected to increase while area under cereals and cotton is expected to decrease as compared to theprevious year. Area and production of principal kharifcrops is given in Table 3.

Exploratory data analysis (EDA): is the crucial process of doing preliminary analyses on data in order to find patterns, identify anomalies, test hypotheses, and verify presumptions with the aid of summary statistics and graphical representations.

Data cleaning: is the process of eliminating or changing data that is inaccurate, lacking, unnecessary, duplicated, or formatted incorrectly in order to prepare it for analysis. When working with categorical data, encoding is a necessary pre-processing step for machine learning algorithms.

Encoding: It is a required pre-processing step when

working with categorical data for machine learning algorithms.

Feature scaling: is a method for uniformly distributing the independent features in the data over a predetermined range. It is done as part of the pre-processing of the data to deal with extremely variable magnitudes, values, or units.

Data Partitioning: The Entire dataset is partitioned into 2 parts: for example, say, 75% of the dataset is used for training the model and 25% of the data is set aside to test the model.

The suggested solution to this problem is shown in Fig. 3 below. Basically, there are two modules:

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 1 Jan. 2023, pp: 1104-1109 www.ijaem.net ISSN: 2395-5252

The Crop Prediction Module is the first onebased on values for temperature, humidity and rainfall; the second is the Crop Yield Prediction Module, which forecasts crop production based on location, season, and area. The OpenWeatherMap API can be used to obtain climate-related data such as temperature and humidity, and the user can provide location and area.

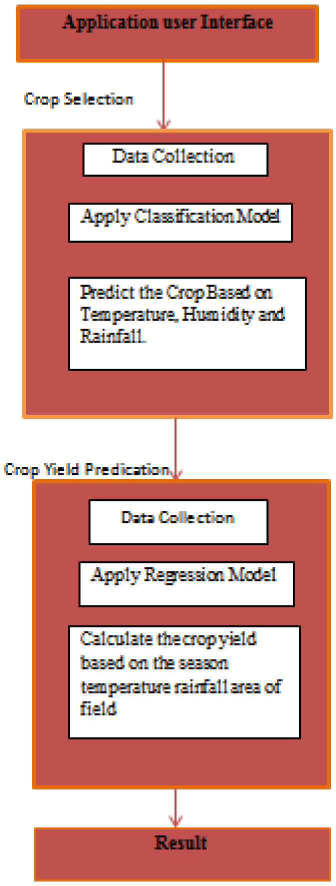


Figure.3 Interface Design

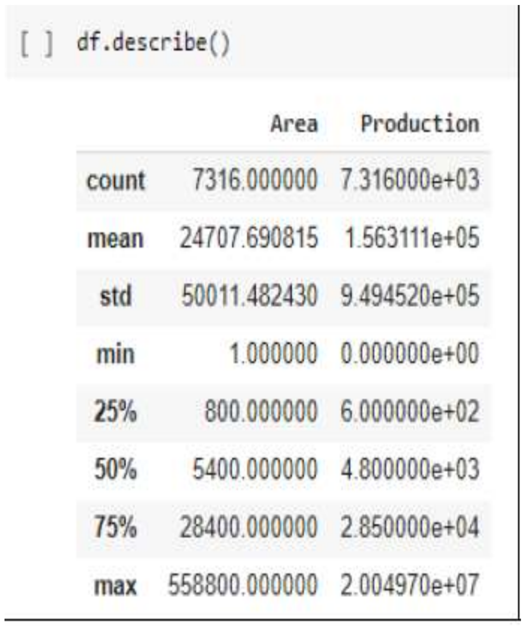
V. IMPLEMENTATION

The following models were created for the first module, which is essentially a multiclass classification problem, and their performance was assessed.
1. KNN
2.Support vector machine
3.Random Forest
4. Naive Bayes
Cohen's Kappa Score served as the evaluation criteria. It is a very effective method that can effectively address problems with both multiple classes and unbalanced classes. It essentially informs you how much better your classifier performs than a classifier that makes random predictions based on how frequently each class occurs.

Classification Models	Cohen's Kappa Score
KNN	0.8825
SVM	0.8108
Random Forest	0.9356
Naive Bayes	0.9513

Table 1: Comparison between various classifiers

The classifier with the highest Cohen's Kappa value is the Naive Bayes Classification Model. So, for this project, the Naive Bayes classifier has been chosen. The following models were created for the second module, which is essentially a regression problem, and their performance was assessed.
1. Multilinear Regression,
2. Regression using Random Forest
3.Supple Vector Regression.
4.KNN Regression.
Outliers make up more than 30% of the dataset for this module. The following figure demonstrates the notable discrepancy between the min, median, mean, and max values, which made it difficult to standardise the data.



```
[ ] df.describe()
```

	Area	Production
count	7316.000000	7.316000e+03
mean	24707.690815	1.563111e+05
std	50011.482430	9.494520e+05
min	1.000000	0.000000e+00
25%	800.000000	6.000000e+02
50%	5400.000000	4.800000e+03
75%	28400.000000	2.850000e+04
max	558800.000000	2.004970e+07

Figure 4.Five Numbers Summary

Hence we divided the dataset into two sets based on the conditions that Area <= 24000 and Area > 24000. And then built the models for both thedatasets. The metrics we used for the evaluation are R-Squared Value and the Mean Squared Error Value(MSE).

[1] Performance of the models for the dataset with observations where Area <= 24000

Regression Model	R-Squared Value	MSE
Random Forest Regressor	0.8285	0.22
Support Vector Regressor	0.0581	1.138
Multiple Linear Regressor	0.1590	1.0168
KNN Regressor	0.7039	0.3580

Table 2: Comparison between various regressors

[2] Performance of the models for the dataset with observations where Area > 24000

Regression Model	R-Squared Value	MSE
Random Forest Regressor	0.9179	0.1011
Support Vector Regressor	0.0119	1.2166
Multiple Linear Regressor	0.2351	0.9418
KNN Regressor	0.8340	0.2043

Table 3: Comparison between various regressors

Random Forest Regression Model gives the highest R-Squared value and least MSE among all the regressors. Hence the Random Forest Regressor has been selected in the project.

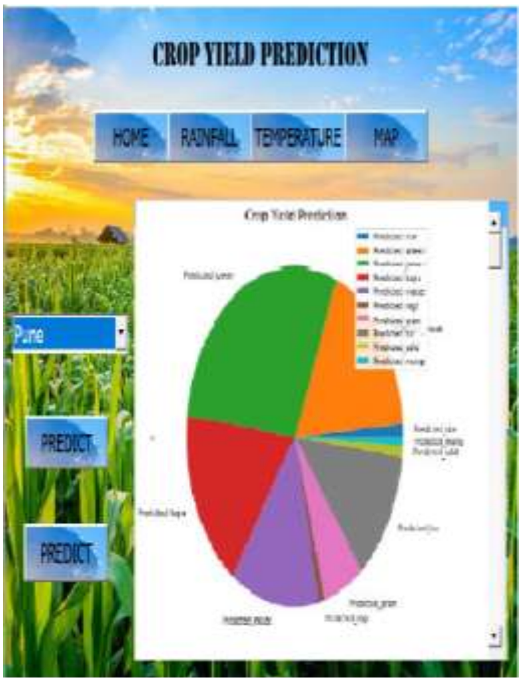


Figure 4 Result

The result shows the appropriate crop based on the Climatic conditions as well as the production in tonnes. Web page also displays the data that user inputs and the weather data.

VI. CONCLUSION

Using machine learning techniques, crop and yield prediction may enhance crop planning decisions. The Naive Bayes Classification Model's Cohen's Kappa score for the Crop Prediction Module is approximately 95%. The Random Forest Regression Model's R-Squared value for the Crop Yield Prediction Module is greater than 81 percent. Future crop and yield forecasts would be accurate with accurate climate parameter forecasts and improved crop historical data.

In addition, the developed webpage is user-friendly and can be made more informative by providing additional helpful information such as fertilizers, intercropping, and other topics. to the client.

REFERENCES

[1]. RashmiPriya, Dharavath Ramesh.2018." Crop Prediction on the Region Belts of India: A Naïve Bayes MapReduce Precision Agricultural Model".

[2]. Igor Oliveira, Renato L. F. Cunha, Bruno Silva, Marco A. S. Netto.2018." A Scalable Machine Learning System for PreSeason Agriculture Yield Forecast."

[3]. Renato L. F. Cunha, Bruno Silva, Marco A. S. Nett. A Scalable Machine Learning

System for Pre-Season Agriculture Yield Forecast. 2018 IEEE 14th International Conference on e-Science (e-Science)
[4]. "data.gov.in", [online] Available: https://data.gov.in/.
[5]. Aruvansh Nigam, SakshamGarg, ArchitAgrawal, Parul AgrawalJaypee Institute of Information Technology, India. Crop Yield Prediction Using Machine Learning Algorithms