SMART FARMING USING MACHINE LEARNING

Article · May 2021

CITATIONS
2

READS
7,793

4 authors, including:

Shubham Pawar
Veermata Jijabai Technological Institute
1 PUBLICATION **2** CITATIONS
SEE PROFILE

Sumit Dere
Veermata Jijabai Technological Institute
1 PUBLICATION **2** CITATIONS
SEE PROFILE

Harshvardhan Kamble
Veermata Jijabai Technological Institute
1 PUBLICATION **2** CITATIONS
SEE PROFILE

SMART FARMING USING MACHINE LEARNING

Shubham Pawar[1], Sumit Dere[1], Ashitosh Akangire[1] , Harshvardhan Kamble[1],

Prof. Seema Shrawne[2]

[1]B.Tech Student, Dept of Computer Engineering & IT , VJTI College, Mumbai, Maharashtra, India.
[2]Assistant Professor, Dept of Computer Engineering & IT, VJTI College, Mumbai, Maharashtra, India.

Abstract — Agriculture plays an important role in Indian economy. But now-a-days, agriculture in India is undergoing a structural change leading to a crisis situation. The only remedy to the crisis is to do all that is possible to make agriculture a profitable enterprise and attract the farmers to continue the crop production activities.
As an effort towards this direction, this research paper would help the farmers in making appropriate decisions regarding the cultivations with the help of machine learning. This paper focuses on predicting the appropriate crop based on the climatic situations and the yield of the crop based on the historic data by using supervised machine learning algorithms. In addition, a web application has been developed.

I. INTRODUCTION

Agriculture is the backbone of the Indian economy. But agriculture in India is undergoing a structural change leading to a crisis situation. The relative contribution of agriculture to the GDP has been declining
over time steadily. It is alarming that India is moving from being a self-reliant nation of food to a net importer of food. All these trends indicate that the agricultural sector in India is facing a crisis today.

It is argued that the consequence of the agricultural crisis in India is very vast and likely to hit all the other sectors and the national economy in several ways. The only remedy to the crisis is to do all that is possible to make agriculture a profitable enterprise and attract the farmers to continue the crop production activities.

In the past farmers used to predict their yield from previous year yield experiences. Thus, for this kind of data analytics in crop prediction, there are different techniques or algorithms, and with the help of those algorithms, we can predict crop yield. Nowadays, modern people don't have awareness about the cultivation of the crops at the right time and at the right place. By analyzing all the issues and problems like weather, temperature, and several factors, there is no proper solution and technologies to overcome the situation faced. Accurate information about history of crop yield is an important thing for making decisions related to agricultural risk management.Therefore, this paper proposes an idea to predict the crop and yield of the crop based on the climatic conditions and historic data related to the crop .The farmer will check the

production of the crop as per the acre ,before cultivating onto the field.

The main objectives are:
a. To use machine learning techniques to predict crop and yield of the crop.
b. To analyze the data properly and to process the data to get better predictions.
c. To improve the performance of machine learning models.
d. To build an easy to use web application.

II. PROBLEM STATEMENT

Crop selection and crop yield prediction are important agricultural problems. The aim of this project is to predict suitable crop based on the given climate parameters and location and also to predict the yield of that crop based on the season and area of the field using machine learning algorithms.

III. LITERATURE REVIEW

In [1] Predicting the yield of the crop using a machine learning algorithm. International Journal of Engineering Science Research Technology. This paper focuses on predicting the yield of the crop based on the existing data by using the Random Forest algorithm. Real data of Tamil Nadu were used for building the model.
http://www.ijesrt.com/issues%20pdf%20file/Archive-2018/April-2018/1.pdf

In [2]. 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.
https://ieeexplore.ieee.org/document/6921718

In [3]. 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.
https://pdfs.semanticscholar.org/3376/e91c3a77a547ce51cfe4a2e68ea6f35ffe63.pdf

In [4]. Random Forests for Global and Regional Crop Yield Predictions. institute on the Environment, University of Minnesota, 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.
https://www.researchgate.net/publication/303794965_Random_Forests_for_Global_and_Regional_Crop_Yield_Predictions

In[5] 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.

In[6] Smart Farming Prediction Using Machine learning
The paper is about using machine learning with various environmental factors like soil, pressure, weather, crop type to predict the maximized profitable crop to grow. The paper mainly focuses on the algorithms used to predict crop yield, crop cost predictions.
https://www.ijitee.org/papers/Smart_Farming_Prediction_Using_Machine_Learning

In [7]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.
https://ieeexplore.ieee.org/document/8554948

IV. ANALYSIS AND DESIGN

To implement the system, we decided to focus on Maharashtra State only in India. Historical data about the crop and the climate at the district level was needed to implement the system. This data has been gathered from the government website www.data.gov.in which includes State, District, Season, Crop, Area and Production. The data about the climate conditions suitable for the particular crops has been collected from the Kaggle which includes Temperature, Humidity, Soil pH, Rainfall and class label is the Crop.

Following figures are the snapshots of the datasets that have been used for this project.

```
= 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

Fig. 1. Dataset For Crop Prediction Problem

```
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

Fig. 2. Dataset For Crop Yield Prediction Problem

Exploratory Data Analysis: It refers to the critical process of performing initial investigations on data so as to discover patterns,to spot anomalies,to test hypotheses and to check assumptions with the help of summary statistics and graphical representations.

Data Cleaning: It is the process of preparing data for analysis by removing or modifying data that is incorrect, incomplete, irrelevant, duplicated, or improperly formatted.

Encoding: It is a required pre-processing step when working with categorical data for machine learning algorithms.

Feature Scaling: It is a technique to standardize the independent features present in the data in a fixed range. It is performed during the data pre-processing to handle highly varying magnitudes or 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 Following Fig. 3 shows the proposed approach to this problem. There are basically two modules: the first one is the Crop prediction module which predicts the most appropriate crop based on temperature, humidity, soil pH and rainfall values and second one is the Crop yield prediction module which predicts the production of the crop based on area, season and location. The climate related data like temperature, humidity etc. can be acquired through OpenWeatherMap API and location, soil pH and area can be input through the user.

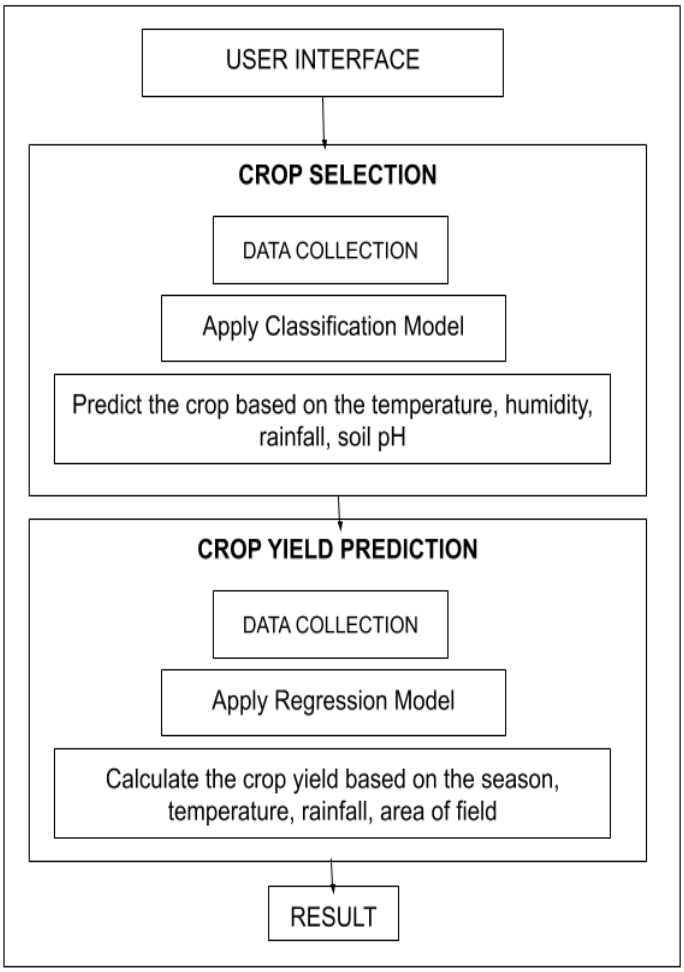

Fig. 3. Proposed Approach

V. IMPLEMENTATION

For the first module which is basically a multiclass classification problem, we built the following models and evaluated their performance.

1. KNN
2. Support vector Machine
3. Random Forest
4. Naive Bayes

The metric we used for the evaluation is Cohen's Kappa Score.. It is a very good measure that can handle very well both multi-class and imbalanced class problems.

It basically tells you how much better your classifier is performing over the performance of a classifier that simply guesses at random according to the frequency of each class.

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

Naive Bayes Classification Model gives the highest Cohen's Kappa score among all the classifiers. Hence the Naive Bayes classifier has been selected in the project.

For the second module which is basically a regression problem, we built the following models and evaluated their performance.

1. Multi-linear Regression
2. Random Forest Regression
3. Support Vector Regression
4. KNN Regression

The dataset for this module contains outliers more than 30% of the dataset. The following figure shows the significant difference between min, median, mean and max values due to which standardizing the data was a big problem.

```
[ ] 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

Fig. 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 the datasets. 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.

To develop an user-friendly web application we used Flask. Flask is an API of Python and it is based on the WSGI toolkit and Jinja2 template engine.

User selects the location, puts the soil ph value, gets from ph meter and then puts the area which is in acres.
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.

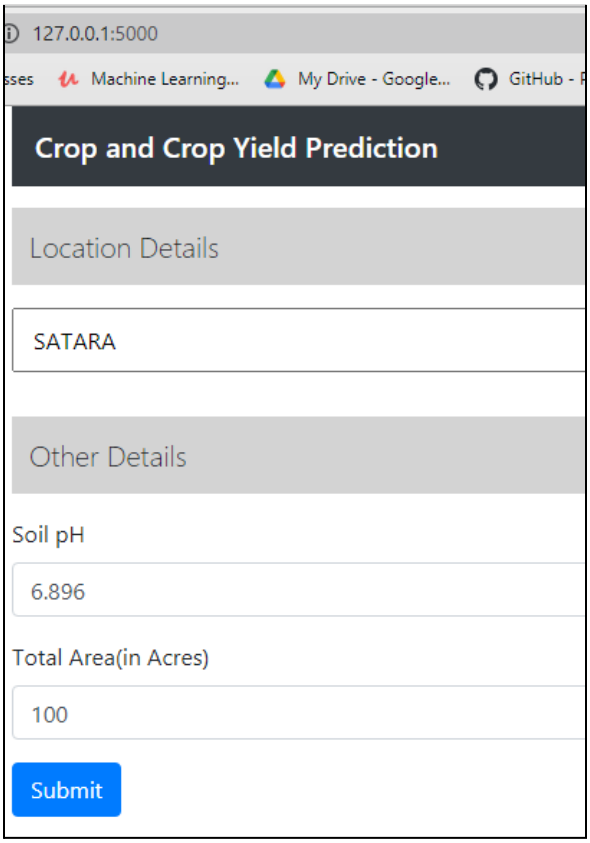

Fig. 5. User Inputs The Data

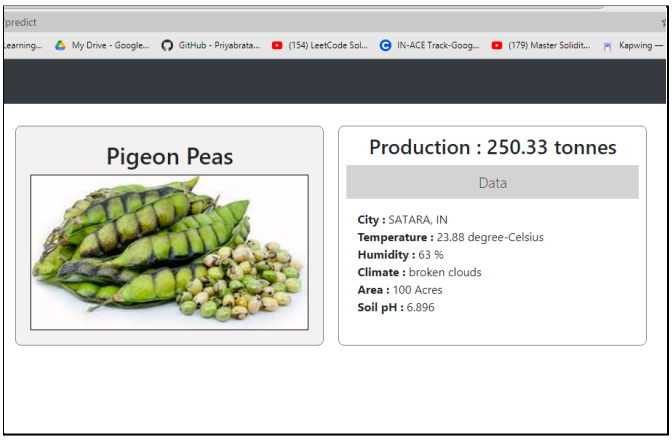

Fig. 6. Predictions

VI. CONCLUSION AND FUTURE SCOPE

Crop and yield of the crop prediction using intelligent machine learning techniques may improve the crop planning decisions. For the Crop Prediction Module, the Cohen's Kappa score we got for the Naive Bayes Classification Model is about 95%. For the Crop Yield Prediction Module, the R-Squared value we got for the Random Forest Regression Model is more than 81%.
Accurate forecasts of the climate parameters and better historic data of the crop would result in accurate crop and its yield forecast in the future. Also, the developed webpage is user friendly and can be made more informative by providing additional useful information like intercropping, fertilizers etc. to the user. We can create more interactive User Interface by adding chatbots and speech recognition systems.

REFERENCES

[1] P.Priya, U.Muthaiah. Balamurugan.Predicting the yield of the crop using a machine learning algorithm. International Journal of Engineering Science Research Technology.

[2]. S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh.Machine learning approach for forecasting crop yield based on climatic parameters. International Conference on Computer Communication and Informatics (ICCCI)

[3]D Ramesh, B Vishnu Vardhan. Analysis Of Crop Yield Prediction Using Data Mining Techniques. IJRET: International Journal of Research in Engineering and Technology

[4]. Jig Han Jeong, Jonathan P. Resop, Nathaniel D. Mueller, David H. Fleisher, Kyungdahm Yun, Ethan E. Butler, Soo-Hyung Kim. Random Forests for Global and Regional Crop Yield Predictions. Institute on the Environment, University of Minnesota, St. Paul, MN 55108, United States of America.

[5]M. Kalimatthu, M. Kishore P.Vaishnavi
Crop Prediction using Machine Learning naive Bayes Gaussian classifier with boosting algorithm.

[6].. Rashmi Priya, Dharavath Ramesh.2018.” Crop Prediction on the Region Belts of India: A Naïve Bayes MapReduce Precision Agricultural Model”.

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

[8]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)

[9] "data.gov.in", [online] Available: https://data.gov.in/.

[10]Aruvansh Nigam, Saksham Garg, Archit Agrawal, Parul Agrawal Jaypee Institute of Information Technology, India. Crop Yield Prediction Using Machine Learning Algorithms

[11] Ramesh Medar, Vijay S. Rajpurohit, Shweta Shweta KLS GIT, Belagavi. Crop Yield Prediction using Machine Learning Techniques

[12] ‘District Wise rainfall data’ Available: http://hydro.imd.gov.in/hydrometweb/(S(ctep2djbwexxp0qr1knt3a45))/DistrictRaifall.aspx

[13] Suvidha Jambekar, Shikha Nema, Zia Saquib. dept.of Electronics and Communication Usha Mittal Institute of Technology Mumbai, Maharashtra, India. Prediction of Crop Production in India Using Data Mining Techniques

[14]Predict Crop Production in India Using Machine Learning Technique: A Survey
Bhavna Sharma, Pratap Singh Yadav, Sunita Yadav

[15]Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique. Rakesh Kumar, M.P. Singh, Prabhat Kumar, J.P. Singh

[16]Machine learning approach for forecasting crop yield based on climatic parameters S. Veenadhari, Dr. Bharat Mishra, Dr. C. D. Singh

[17]Crop yield prediction: two-tiered machine learning model approach Ayush Kapoor, Mrityunjay V. Latte, Sushila Sidnal

[18]Crop Yield Analysis Using Machine Learning Algorithms. Fatin Farhan Haque, Ahmed Abdelgawad, Venkata Prasanth Yanambaka, Kumar Yelamarthi College of Engineering and Technology Central Michigan University, ET100 Mount Pleasant, MI 48858, USA