International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 10 Issue 10 ǁ October 2022 ǁ PP. 558-564

Crop Yield Prediction Based On Indian Agriculture

Dr. Aziz Makandar[1], *Department of Computer Science, Karnataka State Akkmahadevi Women's University Vijayapura.*
Ms. Vrandashree D Koti[2],*PG Scholar, Karnataka State Akkamahadevi Women's University Vijayapura.*

Abstract
India is an Agriculture based economy whose most of the GDP comes from farming. The motivation of this project comes from the increasing suicide rates in farmers which may be due to low harvest in crops. Climate and other environmental changes have become a major threat in the agriculture field. Machine learning is an essential approach for achieving practical and effective solutions for this problem. Predicting yield of the crop from historical available data like weather, soil, rainfall parameters and historic crop yield. We achieved this using the machine learning algorithm. We did a comparative study of various machine learning algorithms, i.e., ANN, K Nearest Neighbour, Random Forest, SVM and Linear Regression and chose Random Forest Algorithm which gave an accuracy of95%.
In this project a mobile application has been developed which predicts the crop yield in general and also for a particular crop. Along with that, it also suggests the user if it is the right time to use the fertilizer ornot.
The impact of climate change in India, most of theagriculturalcropsarebeingbadlyaffectedintermsoftheirperformance over a period of thelast two decades. Predicting thecrop yield in advance of its harvest would help the policy makersand farmers for taking appropriate measures for marketing andstorage. This project will help the farmers to know the yield oftheir crop before cultivating onto the agricultural field and thushelp them to make the appropriate decisions. It attempts to solvethe issue by building aprototype of an interactivepredictionsystem. Implementation of such a system with an easy-to-use webbased graphic user interface and the machine learning algorithmwill be carried out. The results of the prediction will be madeavailable to the farmer. Thus, for such kind of data analytics incrop prediction, there are different techniques or algorithms, andwiththehelpofthosealgorithmswecanpredictcropyield.Randomforestalgorithmisused.
Keywords: *Agriculture, Machine Learning, crop prediction, Supervised Algorithms, Crop yield , Data Mining.*

Date of Submission: 10-10-2022 Date of acceptance: 22-10-2022

I. Introduction

India is ranked 2nd worldwide in farm output [9]. Agriculture and allied sectors like forestry and fisheries accounted for 16.6 percent of the GDP 2009, about 50 percent of the overall workforce [10]. The monetary contribution of agriculture to India's GDP is regularly declining. The crop yield of plants relies on different factors like on climatic, geographical, organic, political and financial elements. For farmers, it is difficult when there is more than one crop to grow especially when the market prices are unknown to them. Citing the Wikipedia statistics, the farmer suicide rate in India has ranged between 1.4 and 1.8 per 100000 total population, over a 10-year period through 2005. While 2014 saw 5650 farmer suicides, the figure crossed 8000 in 2015 [11].

In recent times, it has become inevitable to use technology to create awareness about cultivation. The seasonal climatic conditions are also being changed against the fundamental assets like soil, water and air which lead to insecurity of food. In a scenario, crop yield rate is falling short of meeting the demand consistently and there is a need for a smart system which can solve the problem of decreasing crop yield. Therefore, to eliminate this problem, we propose a system which will provide crop selection based on economic and environmental factors to reap the maximum yield out of it for the farmers which will sequentially help meet the elevating demands for the food supplies in the country. The proposed system uses machine learning to make the predictions. The system will provide crop yield and crop selection based on weather attributes suitable for the crop to get the maximum yield out of it for the farmers. The system makes predictions of the productions of crops by studying the factors such as rainfall, temperature, area (in hectares),season, etc. The system also helps in suggesting whether a particular time is the right one to use fertilizers.

There aremultiplewaysto increaseandimprovethecropyieldandthequalityofthecrops.Dataminingis alsousefulforpredictingcropyieldproduction.Themainobjectivesare :

a) Tousemachinelearningtechniquestopredictcropyield.
b) To provideeasytouseUserInterface.
c) Toincreasetheaccuracyofcropyieldprediction.
d) Toanalysedifferentclimaticparameters(cloudcover,rainfall,temperature).

Objectives:

Collect the weather data, crop yield data, soil type data and the rainfall data and merge these datasets in a structured form and clean the data. Data Cleaning is done to remove inaccurate, incomplete and unreasonable data that increases the quality of the data and hence the overallproductivity.

Scope of theProject *:*

Integrating farming and machine learning, we can lead to further advancements in agriculture by maximizing yield and optimizing the use of resources involved. Previous year's production data is an essential element for predicting the current yield. The goal of this project is to help the farmers by combining agriculture and technology. The end result is an application that is available on the web as well as mobile.

II. Methodology:

Data is a very important part of any Machine Learning System. Toimplement the system, we decided to focus on Maharashtra State inIndia. As the climate changes from place to place, it was necessary togetdataatdistrictlevel.Historicaldataaboutthecropandtheclimateof a particular region was needed to implement the system. This datawas gathered from different government websites.

Crop production is completely dependent upon geographical factors such as soil chemical composition, rainfall, terrain, soil type, temperature etc. These factors play a major role in increasing crop yield. Also, market conditions affect the crop(s) to be grown to gain maximum benefit. We need to consider all the factors altogether to predict the yield.

Hence, using Machine Learning techniques in the agricultural field, we build a system that uses machine learning to make predictions of the production of crops by studying the factors such as rainfall, temperature, area, season,etc.

Table 1: LIST OF ABBREVIATIONS

ANN	Artificial Neural Networks
KNN	K Nearest Neighbors
ML	Machine Learning
SRS	System Requirements Specification
SVM	Support Vector Machine

Proposed System

Our proposed system is a mobile application which predicts name of the crop as well as calculate its corresponding yield. Name of thecrop isdetermined by several features like temperature, humidity, wind-speed, rainfall etc. and yield is determined by the area and production. In this paper, Random Forest classifier is used for prediction. It will attain the crop prediction with best accurate values.

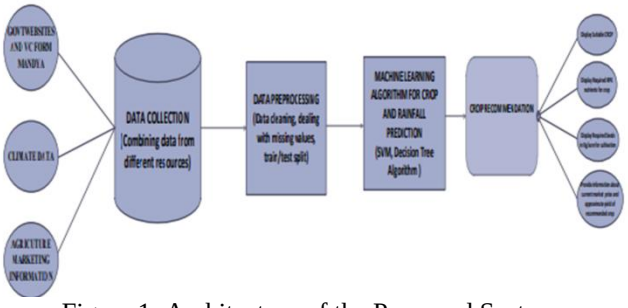

Figure 1: Architecture of the Proposed System

System Architecture: Architecture diagrams can help system designers and developers visualize the high-level, overall structure of their system or application for the purpose of ensuring the system meets their users' needs. They can also be used to describe patterns that are used throughout the design

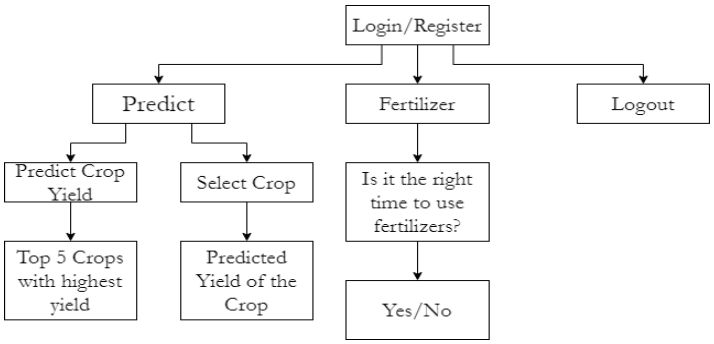

Figure2: System Architecture

The above figure shows the System Architecture. The first step is to login or register to the application. At the next step, three options are available .i.e., Predict, Fertilizer and Logout. The user may select one of the three options and proceed further. Under Predict, the system offers two options that depend on whether the user knows what to plant already or is yet to decide the crop. The inputs are taken from the user in either case and the predicted value is given to the user. When the Fertilizer Module is selected, the user gets a pop up message that says whether or not they can use the fertilizer and it may or may not rain for the next 15 days. Last is the Logout that logs the user out and takes them back to the login/RegisterPage.

FLOWCHART :

IT IS THE GRAPHICAL REPRESENTATION OF THE SYSTEM DESIGNED BY US TO PREDICT THE CROP YIELD BASED ON WEATHER CINDITIONS.

Step 1: The user logs in to thesystem.

Step 2: If login is successful, the location of the user istracked.

Step 3: The system now provides user with the following twopaths:

Prediction Module: The user can choose to know the prediction of a particular cropor know the list of crops with their correspondingproductions.

Yield Prediction: The user needs to provide crop, soil type and area as inputs.The system returns the production of the cropgiven.

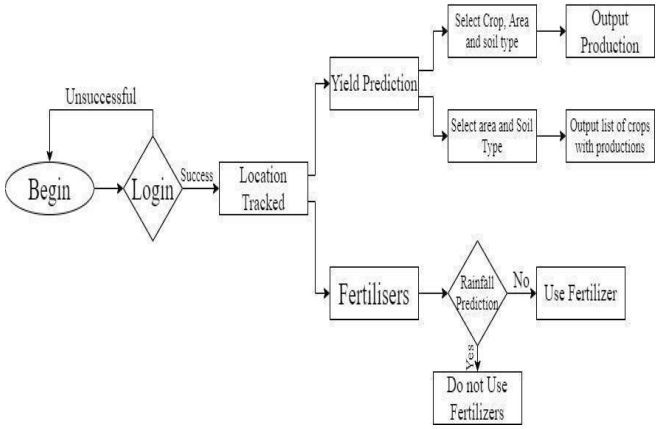

Figure 3: Flowchart of Crop Yield Prediction System

Crops Prediction: The user needs to provide soil type and area as inputs.The system returns a list of crops along with their productionvalues.

Fertilizer Module: The user can choose this module to know if it is the right timeto use the fertilizer. This is done by predicting the rainfall for the next 15 days. The systemreturns yesif it islikelytorain else no

Ifit returns yes the user is suggestednot to usethe fertilizer.

Else the user is suggested to use thefertilizer.

Datasets : Machine Learning depends heavily on data. It‘s the most crucial aspect that makes algorithm training possible.

	Crop	SoilType
1	Crop	SoilType
2	Maize	Sandy
3	Arhar/Tur	Loamy
4	Bajra	Black
5	Gram	Loamy
6	Jowar	Loamy
7	Moong(Green Gram)	Loamy
8	Pulses total	Loamy
9	Ragi	Sandy
10	Rice	Loamy

Figure 4:Soil and Crop data sample

The first step is Data Collection. For this project, we
require two datasets. One for modeling the yield prediction algorithm and other for predicting weather .i.e. Average Rainfall and Average Temperature. These two parameters are predicted so as to be used as inputs for predicting the crop yield. The sources of our datasets are:
1)_https://en.tutiempo.net/'for weather data and 2)_https://www.kaggle.com/srinivas1/agricuture-crops-production-in-india' for crop yield data.

	Year	Season	Avg Rainfall(mm)	Avg Temperature
1	Year	Season	Avg Rainfall(mm)	Avg Temperature
2	1997	Rabi	42.35	27.7
3	1998	Rabi	46.2	27.8
4	1999	Rabi	44.4	27.7
5	2000	Rabi	15.42	27.6
6	2001	Rabi	34.02	27.3
7	2002	Rabi	10.97	27.7
8	2003	Rabi	8.47	27.5
9	2004	Rabi	12.57	27
10	2005	Rabi	16.57	27.5

Figure 5: Rain and Temperature data sample

The yield prediction module dataset requires the following columns: State, District, Crop, Season, Average Temperature, Average Rainfall, Soil Type, Area and Production as these are the major factors that crops depend on. _Production' is the dependent variable or the class variable. There are eight independent variables and 1 dependent variable. We achieved this by merging the datasets. Thedatasets were merged taking the location as the common attribute in both. We are considering only two states here, Maharashtra & Karnataka as the suicide rates in farmers in these two States were found to be veryhigh.

Exploratory DataAnalysis : EDA can give us the following--

- Previewdata
- Check total number of entries and column types using in built functions. It is a good practice to know the columns and their corresponding datatypes.
- Check any nullvalues.
- Check duplicateentries.
- Plot distribution of numeric data (univariate and pairwise jointdistribution)
- Plot count distribution of categorical data.

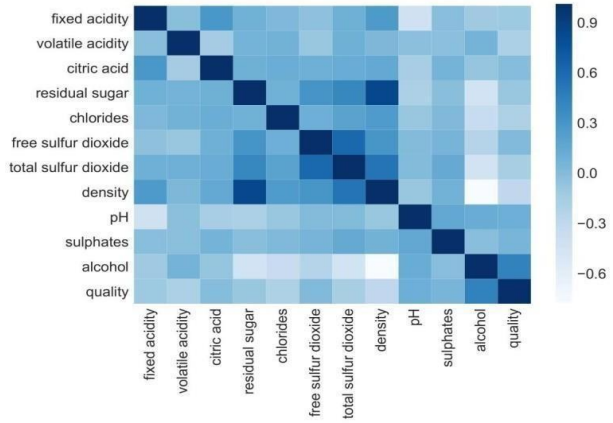

Figure 6: Correlation Matrix Example

Figure 6 shows the simple code written in python to perform the initial steps in EDA .i.e., finding the number of columns, the features, the missing values.
Dark shades represent positive correlation while lighter shades represent negative correlation. Hence, we can make the following inferences from the above example:
Here wecan infer that –density has strongpositive correlation with –residual sugar whereasithas strongnegative correlation with –alcohol.
Freesulphurdioxide and –citricacid have almost no correlation with –quality.

```
df_data.describe()
```

	Year	Area	Production	Avg Rainfall(mm)	Avg Temperature
count	11739.000000	11739.000000	1.161300e+04	11739.000000	10395.000000
mean	2005.280262	26318.246443	5.249072e+04	128.414517	24.936479
std	5.029136	60475.364022	4.157544e+05	97.293491	3.804528
min	1997.000000	1.000000	0.000000e+00	1.100000	19.600000
25%	2001.000000	500.000000	2.890000e+02	16.570000	22.300000
50%	2005.000000	3500.000000	2.400000e+03	166.200000	23.200000
75%	2010.000000	24200.000000	1.990000e+04	208.950000	27.600000
max	2014.000000	726300.000000	1.600010e+07	315.100000	39.900000

Figure 7: Output of describe function

Figure 7 shows how each attribute is related to the other .i.e., the correlation matrix.

EDA Performed:
EDA Code Shows the details of each attribute in tabular form which helps in getting a deeper insight of the attributes.

```
print("Columns  : " , df_data.shape[1])
print("\nFeatures : \n" ,  df_data.columns.tolist())
print("\nMissing values :  ", df_data.isnull().sum().values.sum())
print("\nUnique values :  \n", df_data.nunique())
print("\nInfo : \n")
print(df_data.info())
```

Figure 8: EDA Code

Algorithm Used :
Machine Learning offers a wide range of algorithms to choose from. These are usually divided into classification, regression, clustering and association.

Table 2: Algorithm and Accuracy

Algorithm	**Accuracy**
Artificial Neural Networks (ANN)	86%
Support Vector Machine (SVM)	75%
Multivariate linear Regression	60%
Random Forest	95%
K Nearest Neighbours (KNN)	90%

1)ANN vs. Random Forest: Random Forest is less computationally expensive and does not require a GPU to finish training. A random forest can give you a different interpretation of a decision tree but with better performanceklist exists which is a variety of malware URL data.
2)SVMvs.RandomForest:RandomforestsareprobablyTHE−worry-free|approach.There are no real hyper parameters to tune (maybe except for the number of trees; typically, the more trees we have the better).
3)LinearRegressionvs.RandomForest:Randomforestsveryoftenoutperformlinear regression. Random forests fit data better from the get-go without transforms. They‘re more forgiving in almost every way. You don‘t need to scale your data, you don‘t need to do any monotonic transformations (log, etc.).
4)KNN vs. Random Forest: Random Forest is faster due to KNN‘s expensive real time execution. _K‘ should be wisely selected while there is no such decision to be made in Random Forest. KNN has large computation cost during runtime if sample size islarge.

IMPLEMENTATION
The implementation of the project was divided into two .i.e. crop yield prediction and rainfall prediction (for fertilizers module).

a) Crop YieldPrediction
This module returns the predicted production of crops based on the user's input. If the user wants to know the production of a particular crop, the system takes the crop as the input as well. Else, it returns a list of crops along with their production as output.
These are the following steps of the algorithm implemented:
Step 1: Choose the functionality i.e., crop prediction or yieldprediction.
Step 2: If the user chooses croppprediction:-
Take soil type and area asinputs.
These values are given as input to the random forest implementation in the backend and thecorresponding predictions arereturned.
The algorithm returns a list of crops along with their productionpredicted.
Step 3: If the user chooses yieldprediction:-
Take crop, soil type and area asinputs.
These values are given as input to the random forest implementation in the backend and the corresponding crop yield prediction isreturned.
The algorithm returns the predicted production of the givencrop.

b) FertilizersModule
This module is used to suggest the farmer on usage of fertilizer based on the rainfall in next few days. To predict the rainfall for the next 15 days we are using an API service provided by _OpenWeather‘. If it is likely to rain we suggest the farmer not to use the fertilizer.
These are the following steps of the algorithm implemented:
Step 1: On selection of this module, API call is made to the _OpenWeather‘Services.
Step 2: The rainfall for the next 14 days is read from the result of the APIcall.
c) ExperimentalImplementation
The implementation of the system can be divided into two, i.e., frontend and backend implementation.

The frontend is implemented using the ionic development tools. Ionic Framework is an open source UI toolkit for building performance, high-quality mobile and desktop apps using web technologies — HTML, CSS, and JavaScript — with integrations for popular frameworks like Angular and React. Ionic Framework focuses on the frontend UX and UI interaction of an app — UI controls, interactions, gestures, animations. It integrates with other libraries or frameworks, such as Angular, React, or Vue and thus can be used on any platform.

III. CONCLUSION AND FUTURE SCOPE

This system is proposed to deal with the increasing rate of farmer suicides and to help them to grow financially stronger. The Crop Recommender system helps the farmers to predict the yield of a given crop and also helps them to decide which crop to grow. Moreover, it also tells the user the right time to use thefertilizer.

Appropriate datasets were collected, studied and trained using machine learning tools. The system tracks the user‘s location and fetches needed information from the backend based on the location. Thus, the user needs to provide limited information like the soil type andarea.This system contributes to the field of agriculture. One of the most important and novel contributions of the system is suggesting the user the right time to use the fertilizer, this is done by predicting the weather of the next 14 days. Also, the system provides a list of crops with their productions based on the climatic conditions.

The future work is focused on providing the sequence of crops to be grown depending on the soil and weather conditions and to update the datasets time to time to produce accurate predictions. The Future Work targets a fully automated system that will do the same. Another functionality that we are trying to implement is to provide the correct fertilizer for the given crop and location. To implement this through study of fertilizers and their relationship with soil and climate is required. We are also aiming to predict the crisis situation in advance like the recent hike of onion prices.

REFERENCES

[1]. https://towardsdatascience.com/machine-learning-basics-part-1-a36d38c7916
[2]. https://healthcare.ai/machine-learning-versus-statistics-use/
[3]. https://medium.com/fintechexplained/machine-learning-algorithm-comparison- f14ce372b855
[4]. https://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learing/
[5]. https://scikitlearn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html
[6]. https://openweathermap.org/api/one-call api?gclid=CjwKCAjwtL2BRA_EiwAacX32SzjPujH29VetWiBl-xcyG9Dw17xto02vSwkxdTFeIR18tr5GF- 8vxoC2- kQAvD_BwE
[7]. https://ukdiss.com/examples/rainfall-prediction-machine-learning.php
[8]. https://iopscience.iop.org/article/10.1088/1748-9326/aae159
[9]. https://en.wikipedia.org/wiki/Agriculture_in_India
[10]. https://www.agricultureinformation.com/forums/threads/agriculture-in-india.101537/
[11]. https://en.wikipedia.org/wiki/Farmers%27_suicides_in_India
[12]. https://towardsdatascience.com/machine-learning-an-introduction-23b84d51e6d0
[13]. Groundnut Crop Yield Prediction Using Machine Learning Techniques by VinitaShah and PrachiShah.Developing regression model to forecast the rice yield at Raipur condition by A Jain, JL Chaudhary, MK Beck and LoveKumar
[14]. Machine learning approach for forecasting crop yield based on climatic parameters by S.Veenadhari, Dr. Bharat Misra& Dr. CD Singh, International Conference on Computer Communication and Informatics (ICCCI -2014), Jan, 2014
[15]. Predicting Yield of the Crop Using Machine Learning Algorithm by P.Priya, U.Muthaiah&M.Balamurugan, International Journal of Engineering Sciences &Research Technology (IJESRT), April,2018