International Journal of Early Childhood Special Education (INT-JECSE)
DOI:10.9756/INTJECSE/V14I6.139 ISSN: 1308-5581 Vol 14, Issue 05 2022

AGRICULTURE-CROP RECOMMENDATION SYSTEM BASED ON PRODUCTIVITY AND SEASON

Dr.Amjan Shaik
Professor
Dept. of.CSE
St. Peter's Engineering college(A)
Hyderabad,TS,India
hod.cse@stpetershyd.com

Karthi Rayala
IV B.Tech
Dept.of.CSE
St. Peter's Engineering college(A)
Hyderabad, TS,India
karthichowdary2228@gmail.com

Ravi Rahul Varma Kosuru
IV B.Tech
Dept.of.CSE
St. Peter's Engineering College(A)
Hyderabad,TS, India
ravirahulvarma125@gmail.com

Divya Solanke
IV B.Tech
Dept.of.CSE
St. Peter's Enginnering College (A)
Hyderabad,TS, India
solankedivya20@gmail.com

***Abstract*— India is an husbandry grounded frugality whose utmost of the GDP comes from husbandry. The provocation of this design comes from the adding self-murder rates in growers which may be due to low crop in crops. Climate and other environmental changes have come a major trouble in the husbandry field. Machine literacy is an essential approach for achieving practical and effective results for this problem. Predicting yield of the crop from literal available data like Rainfall, soil, downfall parameters and major crop yield. We achieved this using the machine learning algorithm. We did a relative study of colorful machine learning algorithms, i.e., ANN, K Nearest Neighbour, Random Forest, SVM and Linear Retrogression and chose Random Forest Algorithm which gave an delicacy of 95. In this design a mobile operation has been developed which predicts the crop yield in general and also for a particular crop. Along with that, it also suggests the stoner if it is the right time to use the toxin or not.**

Keywords—GDP, ANN , KNN , SVM , SRM

I. INTRODUCTION -AGRICULTURE

India is ranked 2nd worldwide in farm output[9]. 16.6% of the GDP in 2009 came from the agricultural and related industries, which employed about 50% of the total workforce[10]. Agriculture's financial contribution to India's GDP is steadily shrinking. Plant crop output is influenced by a variety of factors like on climatic, geographical, organic, political and financial elements.

It is challenging for farmers to raise more than one crop, especially when they are unaware of market values. According to data from Wikipedia, between 2004 and 2005, the rate of farmer suicide in India varied between 1.4 and 1.8 per 100,000 people. While there were 5650 farmer suicides in 2014, the number surpassed 8000 in 2015 [11]. Utilizing technology to spread knowledge about agriculture has become a necessity in recent years. Food insecurity is a result of the seasonal climate changing in opposition to natural resources including soil, water, and air.

There is a need for a clever system that can address the issue of declining agricultural output in a situation where crop yield rate is lowering and needs to be maintained to satisfy demand. Therefore, to eliminate this problem, we propose a system which can provide crop selection supported economic and environmental factors to reap the most yield out of it for the farmers which is able to sequentially help meet the elevating demands for the food supplies within the country. The proposed system uses machine learning to form the predictions. In order to acquire the best yield from the farmers, this system offers crop yields and crop selection support based on meteorological variables appropriate for the crop.

By analyzing variables such as rainfall, temperature, area (measured in hectares), season, and other factors, the method predicts crop yield. The method also helps in deciding whether fertilizers should be used at a particular time. Forecasting crop yields is a crucial agricultural issue. Every farmer is constantly worried about the yield percentage that will be generated and whether it will meet their expectations. In the past, yield prediction was derived by analyzing a farmer's prior experience with a particular crop. Weather, pests, and harvest process layout are the main factors that determine agricultural productivity. Making judgments regarding agricultural risk management depends on having accurate information about crop production history.

II.INFORMATION RESEARCH IN DATABASES

The course mining process involves extracting data from the informative index. It anticipates providing ranchers with precise results. It unearths covert examples. The enormous informative index yields useful information. One of the cycles in Knowledge Revelation in Databases includes this one (KDD). In addition to the KDD cycle, machine learning—which also includes high performance computing—has recently evolved to handle massive volumes of information. The use of machine learning in agriculture is steadily increasing. Machine learning techniques are used in agricultural management, domestic animal management, water management, and soil management.

The use of a Proposal computation is one type of machine inclining procedure. Customized products are offered through e-commerce. In order to provide harvests for plants, this paper uses these suggestion principles in horticulture. Farmers are advised to personalise their agrarian harvests by using basic data analytics on crop datasets.

III.OBJECTIVES

By predicting agricultural yields under certain climatic conditions, this study aims to recommend suitable crops for that site. The subsequent actions are involved.

- Gather information on the weather, crop yield, soil type, and rainfall, combine these facts in an organized manner, 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 overall productivity.
- Performing Exploratory Data Analysis (EDA) that helps in analyzing the complete dataset and summarizing the main characteristics.
- It is used to discover patterns, spot anomalies and to get graphical representations of various attributes. Most importantly, it tells us the importance of each attribute, the way that each attribute is affected by the class attribute as well as other important details.
- By using the crop data under study, create training and testing sets, and then train the model using the training data to predict the crop output given a set of inputs.
- By running the examined dataset through several algorithms and determining each one's error rate and accuracy, you may compare them. The method with the best accuracy and lowest error rate should be chosen.
- Integrate the algorithm into the back end of a system that is implemented as a mobile application.
- To check for accuracy and flaws, test the implemented system.

IV.METHODOLOGY

Since Python is generally recognized as a language for machine learning experimentation, the system uses machine learning to estimate the crop and uses Python as the programming language. Machine learning employs past data and knowledge to build experience and create a trained model by training it with data. The output of this model is then predicted. The classifier's accuracy will improve with a better dataset collection. Regression and classification are two machine learning techniques that appear to perform better than other statistical models[2].

The production of crops is entirely reliant on geographic elements such as soil chemistry, rainfall, terrain, soil type, and temperature. These elements significantly contribute to raising crop output.

Additionally, market factors impact which crop(s) should be planted to reap the greatest rewards. In order to anticipate the yield, we would like to consider all the factors jointly. In order to anticipate the assembly of crops by researching variables like rainfall, temperature, area, season, etc., We developed a system that applies machine learning principles to the agricultural sector.

Unquestionably, Machine learning is one of the most significant and effective technologies in use today. a device for turning knowledge into information is machine learning. Data has multiplied exponentially during the last 50 years [10].

V. PRIMARY PROCESS

i. Data Gathering: Compile the statistics from which the set of rules will be derived.

ii. Data Preparation: Format the information in the best possible way, acting to reduce dimensionality and extract essential functions.

iii. Training: Also known as the correct stage, this is the stage in which the machine learning set of rules really learn by being shown the gathered and prepared data.

iv. Assessment: Try out the version to see how well it works.

v.Tuning: Optimize the model's performance by fine-tuning it.

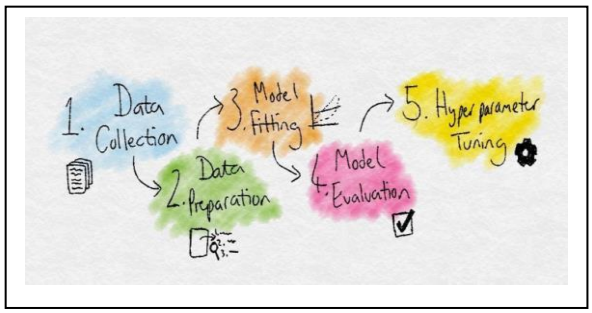


Fig 1: Primary Process

A.DATA COLLECTION

The dataset used on this undertaking is the facts gathered from dependable web sites and merged to obtain the preferred facts set. Our datasets have been reassessed at https://en.tutiempo.net/ for information about the climate, and at https://www.kaggle.com/srinivas1/agriculture-crops-production-in-india. for data on agricultural yield. The names of the crops, production, area, typical temperature, typical rainfall (mm), season, year, and the names of the states and districts are all included. The structured variable, sometimes known as the magnificence variable, is production. Eight unbiased variables and one structured variable are present[18].

A.1EXPLORATORY DATA ANALYSIS

- Display the data
- Check the total number of entries and the types of columns using the built-in functions. It's a good habit to be familiar with the columns and the corresponding data types.
- Look for any null values.
- Look for overlapping entries
- Plot categorical data count distribution
- Plot distribution of numerical data (univariate and pairwise joint distribution).

Table 1: Data Uploaded

State	District	Year	Season	Crop	Area	Production	SoilType	Avg Rainfall	Avg Temperature
Maharashtra	AHMEDNAGAR	1997	Kharif	Arhar/Tur	17600	6300	Loamy	184.63	22.6
Maharashtra	AKOLA	1997	Kharif	Arhar/Tur	81200	64400	Loamy	184.63	22.6
Maharashtra	AMRAVATI	1997	Kharif	Arhar/Tur	83400	61300	Loamy	184.63	22.6
Maharashtra	AURANGABAD	1997	Kharif	Arhar/Tur	37100	3700	Loamy	184.63	22.6
Maharashtra	BEED	1997	Kharif	Arhar/Tur	44200	7200	Loamy	184.63	22.6
Maharashtra	BHANDARA	1997	Kharif	Arhar/Tur	10200	2700	Loamy	184.63	22.6
Maharashtra	BULDHANA	1997	Kharif	Arhar/Tur	63000	29700	Loamy	184.63	22.6
Maharashtra	CHANDRAPUR	1997	Kharif	Arhar/Tur	30300	9700	Loamy	184.63	22.6
Maharashtra	DHULE	1997	Kharif	Arhar/Tur	29400	18500	Loamy	184.63	22.6
Maharashtra	GADCHIROLI	1997	Kharif	Arhar/Tur	2100	600	Loamy	184.63	22.6
Maharashtra	JALGAON	1997	Kharif	Arhar/Tur	24100	7300	Loamy	184.63	22.6
Maharashtra	JALNA	1997	Kharif	Arhar/Tur	37700	5600	Loamy	184.63	22.6
Maharashtra	KOLHAPUR	1997	Kharif	Arhar/Tur	3700	900	Loamy	184.63	22.6
Maharashtra	LATUR	1997	Kharif	Arhar/Tur	65000	2800	Loamy	184.63	22.6
Maharashtra	NAGPUR	1997	Kharif	Arhar/Tur	51900	2500	Loamy	184.63	22.6
Maharashtra	NANDED	1997	Kharif	Arhar/Tur	46000	6000	Loamy	184.63	22.6
Maharashtra	NASHIK	1997	Kharif	Arhar/Tur	8800	4300	Loamy	184.63	22.6
Maharashtra	OSMANABAD	1997	Kharif	Arhar/Tur	69200	2900	Loamy	184.63	22.6

B.DATA PREPROCEESING

Preprocessing is done to transform raw data into a format that can be used by machine learning. A data scientist can use an applied machine learning model to obtain more accurate findings by using structured and clean data. Data formatting, cleansing, and sampling are all part of the method. Here, data pre-processing focuses on finding the attributes with null values or invalid values and finding the relationships between various attributes as well.

Data Pre-processing also helps in finding out the impact of each parameter on the target parameter[1]. Our datasets were preprocessed using the EDA approach. The invalid and null values were all dealt with by removing that record or giving the default value of that particular attribute based on its importance

C.DATA SPLITING

A dataset used for system mastering have to be partitioned in to subsets education and check sets. We break up the dataset into with a break up ratio of eighty% i.e., in a hundred statistics eighty statistics have been part of the education set and closing 20 statistics have been part of the check set.

D. MODEL INSTRUMENTATION

A form of education can proceed after a records scientist has preprocessed the obtained data and divided it into teach and look at. This technique entails ―feeding‖ the set of rules with schooling records. An set of rules will technique records and output a version this is capable of discover a goal new records' value (attribute) a remedy is required to obtain a predictive analysis. The motive of version schooling is to increase a version. We educated our version the use of the random woodland set of rules. On schooling the version it estimates the yield on giving the alternative input attributes from the dataset.

E. EVALUATION AND TESTING OF MODELS

This step's objective is to broaden the most effective version that can quickly and accurately formulate a goal cost. Through version adjustment, an information scientist can achieve this goal. That is version parameter optimization to achieve good algorithm performance.

II. ALGORITHM'S USED

There are many available algorithms for machine learning. These are typically broken down into association,grouping,regression,and classification. supervised learning encompasses techniques for classification and regression, whereas unsupervised learning encompasses clustering and association.[15].

Classification: The scenario is referred to be a classification problem when the output variable is a category, such as "red," "blue," "sickness,"

or "no disease." Taking decision trees as an example.

Regression: The scenario is referred to as a regression problem when the output variable has a real value, such as money or weight. Taking linear regression as an example.

Identifying the underlying categories in the data, such as classifying customers according to their purchasing patterns, is known as clustering. Examples include clustering using "k" means. [12].

In an association rule learning task, you aim to uncover rules that broadly characterise your data, such as "people who buy X also tend to buy Y." Example includes Apriori Algorithm.

Some algorithms may fall under more than one kind. Regression would be the most appropriate kind of algorithm given the project's goal and the problem statement. Numerous algorithms were investigated, and their accuracy and error rates were evaluated, prior to selecting one and working with it further.[3].

This paper also compares the two machine learning algorithms: decision trees and Random Forest.

Decision Tree: Due to the greedy strategy employed by decision tree classifiers, an attribute selected in the first stage cannot be utilized again even though it could improve classification if used in subsequent steps. Additionally, it overfits the training data, producing subpar unseen data. So, ensemble modeling is employed to get over this restriction. Ensemble models mix the outputs of various models. An ensemble model's output typically outperforms the output of any single individual model.

Random Forest: An ensemble classifier called Random Forests makes use of several decision tree models to forecast the outcome. To train each tree, a different subset of the training data is chosen. A forest is a collection of trees, and therefore the trees in random forests are collections of trees since they were trained on subsets that were randomly selected. Problems involving classification and regression can be solved with this. The number of votes from all the trees determines the class, and the average of the results is utilized for regression. The steps they took in this paper are listed below[4].

$$\mathbf{E(T,X)=\Sigma P(c)E(c)}$$

Where

E(T,X)=E(Yield,Season)

P(c)=Winter,Summer,Autumn,Rainy,Spring

E(c)=Area

- Divide the loaded data sets into training and test data sets using a split ratio of 67 percent and 33 percent, respectively [5].
- Once the necessary tuples have been calculated, summarize the data sets using the Mean and Standard Deviation. Calculate the likelihood by comparing the original data sets and the list of summary data.
- The biggest probability generated is used for prediction based on the outcome. By contrasting the derived class value with the test data set, the accuracy can be estimated. The accuracy can be in the range of 0% and 100%..

Another Reason for choosing Random Forest: The data of a particular crop was taken and passed through two algorithms .i.e., Random Forest and another Algorithm that is said to give best results for that crop. The accuracy achieved in both the algorithms were compared. Rice and Groundnut were chosen based on the research papers that were found.

Rice: The paper claims that linear regression is the best approach for predicting rice yield. [14]. Following the execution of both algorithms, we discovered that while Linear Regression continued to retain an accuracy of 90+, there was a significant disparity between the real value and predicted value.

Groundnut: The research article states that KNN has the greatest performance in forecasting groundnut yield [13]. We found little difference between the two algorithms' findings after running them.

VI. AGRICULTURAL RECOMMENDATION TECHNIQUES

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. The application has the following features:

i. Login/Register: To register, a user must enter a username and password of their choice. After registering, individuals can log in to continue using the application and see all of their available options.

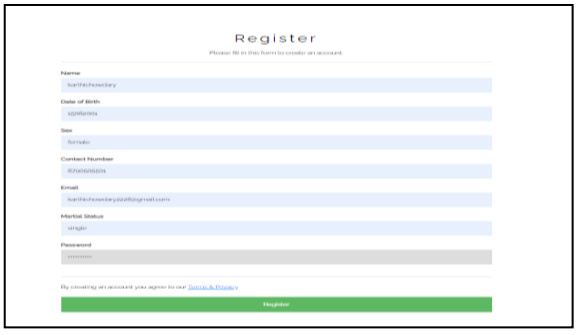


Fig 2: Registration page

ii. Yield Prediction: One of the modules in the application that allows users to view crop production forecasts is called "yield prediction." Here, the user has two options:

_I know what to plant‘: This option is for those users who already have a crop in their mind that they want to grow. When chosen, the user will be given choices of crops that they must select along with other inputs .i.e., Area and the soil type. After processing the inputs, the application will return the predicted yield on the user‘s screen[16].

_Yet to decide the crop‘: This option is when the user is not sure between some crops or has no crop in mind. The user must enter the location and the type of soil. The modeled algorithm once more processes the input at the back end, returning to the user the expected yield.

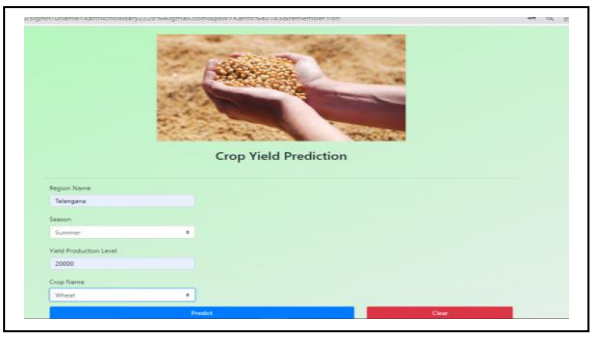


Fig 3: Crop Yield page

iii. Fertiliser: This is the second module available. The functionality that this module provides revolves around whether using fertilisers at a certain point of time would be recommended or not. As farmers mostly use water soluble fertilizers, it is important that it doesn‘t rain for 14- 15 days after they use the fertilisers as they may wash off with rain and the use of the fertilisers will go in vain.

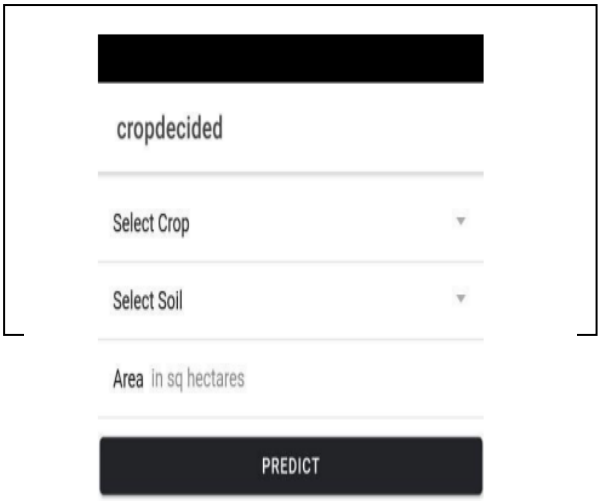


Fig 4: Crop Prediction Page

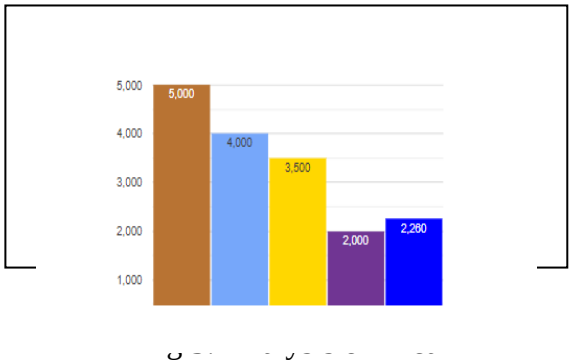
iii. *Sign Out:* Signing out will return the user to the login/register page at the conclusion. Given that farmer suicides are on the rise and that agriculture is one of India's major industries, it is crucial to maintain agricultural sustainability. As a result, it makes a considerable contribution to the wellbeing of the global economy and agriculture.

The majority of the existing systems, according to this paper, are hardware-based, which makes them expensive, difficult to maintain, and unable to produce reliable results. Depending on yield rate and market price, several methods recommend a specific crop sequence.The suggested approach makes an effort to get around these limitations and forecasts crops using structured data. It does not allow for much consideration of maintenance since it is a wholly software solution. Additionally, the accuracy would be higher than with hardware-based solutions because the prediction method takes into account factors like soil composition, soil type, pH value, and meteorological conditions[6].

The results of the study demonstrate that we can accurately anticipate agricultural yield using the Random Forest method, according to the paper's conclusion. With the fewest models, the Random Forest method produces the greatest number of crop yield models. It is suitable for estimating large-scale crop yields in agricultural planning. The dataset used for modeling here includes the climatic factors as well i.e., rainfall and temperature. The author did a comparative study of decision trees and random forest algorithms. But other algorithms were not considered and the dataset includes very few attributes that would not give accurate predictions.

VII.CONCLUSION&DISCUSSION

- This system is projected to wear down the increasing rate of farmer suicides and to assist them to grow financially stronger. The Crop Recommender system aids farmers in forecasting the output of a specific crop and conjointly helps them to determine that crop to grow. Moreover, it conjointly tells the user the proper time to use the plant food.



- Appropriate datasets were collected, studied and trained mistreatment machine learning tools. The system tracks the user‘s location and fetches required data from the backend supported the situation. Thus, the user must give restricted data just like the soil sort and space.
- This system contributes to the sector of agriculture. One in all the foremost vital and novel contributions of the system is suggesting the user the proper time to use the plant food, this can be done by predicting the weather of future fourteen days. Also, the system provides a listing of crops with their productions supported the atmospheric condition[17].

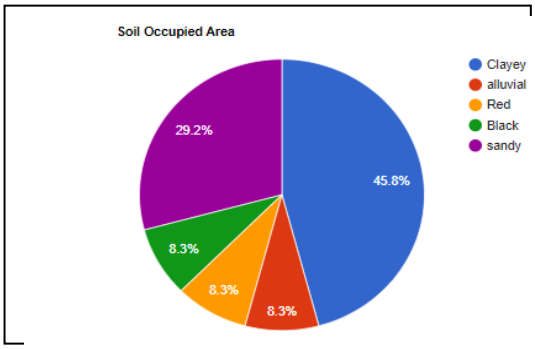


Fig 6: Analysis of Soil occupied Area

- The future work is targeted on providing the sequence of crops to be adult betting on the soil and atmospheric condition and to update the datasets periodically to produce accurate predictions. The long run Work targets a completely machine-controlled system that may do identical. Another practicality that we have a tendency to try to implement is to supply the proper plant food food for the specified crop and location. It is necessary to put this into practise by researching fertilizer's effects on soil and climate. We have a tendency to are planning to predict the crisis scenario earlier just like the recent hike of onion costs.

VIII.ACKNOWLEDGEMENT

We are thankful to St. Peter's Engineering College, Department of CSE for helping us with laboratory and continuing support to prepare this paper in a brighter manner.

VIII.REFERENCES

[1] https://healthcare.ai/machine-learning-versus-statistics-use/

[2] Developing regression model to forecast the rice yield at Raipur condition by A Jain, JL Chaudhary, MK Beck and Love Kumar

[3] https://medium.com/fintechexplained/machine-learning-algorithm-comparison f14ce372b855

[4] A Survey on Crop Prediction using Machine Learning Approach by Sriram Rakshith.K, Dr. Deepak.G, Rajesh M, Sudharshan K S, Vasanth S & Harish Kumar, International Journal for Research in Applied Science & Engineering Technology (IJRASET) Volume 7, Issue IV, Apr 2019

[5] https://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learing/

[6] https://scikit learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

[7] https://ukdiss.com/examples/rainfall-prediction-machine-learning.php

[8] https://www.agricultureinformation.com/forums/threads/agriculture-in-india.101537/

[9] https://iopscience.iop.org/article/10.1088/1748-9326/aae159

[10] https://en.wikipedia.org/wiki/Agriculture_in_India

[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 Vinita Shah and Prachi Shah.

[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] https://openweathermap.org/api/one-call-api?gclid=CjwKCAjwt L2BRA_EiwAacX32 SzjPujH29VetWiBl-xcyG9Dw17xto02vSwkxdTFeIR18tr5GF-8vxoC2- kQAvD_BwE

[16] 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

[17] Heuristic Prediction of Crop Yield using Machine Learning Technique by S. Pavani,

Augusta Sophy Beulet P, International Journal of Engineering and Advanced Technology (IJEAT) Volume-9, December 2019.

[18] https://towardsdatascience.com/machine-learning-basics-part-1-a36d38c7916.