

# Comparative Analysis of Machine Learning Algorithms in The Study of Crop and Crop yield Prediction

S Bharath<sup>1\*</sup>

Department of Computer Science and Engineering  
SJB Institute of Technology  
Bengaluru, India

Yeshwanth S<sup>2</sup>

Department of Computer Science and Engineering  
SJB Institute of Technology  
Bengaluru, India

Yashas B L<sup>3</sup>

Department of Computer Science and Engineering  
SJB Institute of Technology  
Bengaluru, India

Vidyaranya R Javalagi<sup>4</sup>

Department of Computer Science and Engineering  
SJB Institute of Technology  
Bengaluru, India

**Abstract**— Farming analysts demand the requirement for an effective system to play an imperative and significant job in the development of national economy. In a nation like India agribusiness contributes about 20% of national GDP. This expectation will assist the ranchers with choosing relevant harvests for development in their homestead as per the climate gauging components, for example temperature, precipitation and furthermore soil ph. All these authorize of information will be inspected. We will prepare the information with different appropriate ML calculations for making the necessary model. The framework accompanies a model that is pinpoint and impeccable in anticipating crop yield and gives the end client appropriate backing about the harvest that can be become dependent on climatic and soil boundaries of the land which improve to expand the harvest yield and increment rancher turnover.

**Index Terms**—Comparing Algorithms, Crop yield, Machine Learning, Prediction, Crop.

## I. INTRODUCTION

The principle mainstay of any nation's economy will be Agriculture. In a Country like India, which has ever flood request of food because of the rising populace, quantum jump in horticulture area are required to address the issues of the rising populace. Subsequently, the normal harvests are developed and have been utilized by numerous animals, for example, people, creatures and winged creatures. Because of the extravagant innovation individuals are been exhausting on developing counterfeit items that is amalgam items where this prompts an unfortunate way of life. These days, present day individuals don't have awareness about the development of the harvests in an ideal time and at an opportune spot. As a result of these developing strategies the occasional climatic conditions are additionally being changed against the simple resources like soil, water and air which lead to scarcity of food.

By investigating every one of these issues and complexities like climate, temperature and a few components, there is no appropriate arrangement and advancements to defeat the circumstance looked by us. In India there are a few different ways to build the opportune and monetary development in the

field of agribusiness. There are different approaches to increment and improve the harvest yield and the matchless quality of the harvests. Harvest yields are basically hand-off on climate. Utilizing information on yield from the locally accessible dataset, we show that this methodology outflank both old style factual techniques during model preparing. Fastidious models planning climate to edit yields are significant for anticipating effects on farming, yet additionally for anticipating the repercussions of environmental change on partner financial and ecological results, and thusly for alleviation and adjustment strategy.

In equal, AI (ML) strategies have move forward significantly in the course of recent decades. ML is insightfully recognizable from quite a bit of old style measurements, to a great extent since its objectives are unique. It is to a great extent concentrated on guess of results, rather than induction into the idea of the unthinking procedures creating those results. The AI presents a few strategies to characterize rules and examples in enormous informational collections identified with crop yield and have notable anticipating ability. Moreover, it can self-ad lib the prescient model.

## II. LITERATURE SURVEY

Many research had been carried out with intention of predicting high accurate crops based on the types of soil, soil pH, varying temperature and average rainfall across the region. Many models are also built for predicting yield of crop, detecting diseases in crop and usage of organic and non-organic pesticides for the better yield production. Zaminur rahman et al. [1] prepares a dataset by common understanding of each kind of soil has different highlights and various sorts of harvests can be grown on suitable type of soil. In this he proposed the model that can foresee soil arrangement with kind of land and with methods like KNN and SVM suitable crop is suggested. S.Veenadhari et al.[4] proposed a model of predicting crop based on the climatic parameter. Study shows how much climate change can impact agriculture growth in India, proper yield expectation is highly determined by varying

climate change. In [4] a user friendly software tool ‘Crop Advisor’ predicting the influence of climate on crop is discussed which is developed by C4.5 algorithm in the state of Madhya Pradesh. Similar paper [5] by Sonal Jain et al [5] discussed on crop yield on basis of changing weather conditions in India. This model also suggests correct method and proper sowing time for suitable crop, machine learning techniques like recurrent neural network compared with conventional Artificial Neural Network was used in selecting suitable crop. Yunous vagh et al [2] considered average monthly temperature profiles of some zone were utilized to recognize the impact of temperature on crop creation. Research focuses on the crops grown in south western parts of Australia from 2002-2005, evaluation is completed utilizing graphical relationship and information mining relapse procedures. Model predicts that crop yield is higher in the increasing temperature areas. Igor Oliveira et al [3] proposes how yield forecast can be obtained which rely on Normalized Difference Vegetation Index data. This paper describes the system which consolidates satellite determined precipitation and soil property dataset, occasional gauging information from physical models which helps in removing high goal remote detecting information and help ranchers to get ready for the impact of atmosphere in crop cycle. Niketa Gandhi et al. [6] describe the Rice crop yield prediction across the country under various climatic situations. This paper talks about the outcomes acquired by applying SMO classifier in predicting crop. To study this mean absolute error, root mean square error, relative absolute error was calculated on analyzing the values obtained crop is predicted. Rakesh Kumar et al. [7] proposes system which helps in selection of crop can be cultivated in the same piece of land. This paper highlights the use of method name Crop Selection Method to solve selection of type of crop problem and may improve current yield quantity.

### III. PROPOSED SOLUTION

#### A. Narration of the Dataset

The dataset in Table I given below depicts the sample of the dataset that is used the proposed approach. It consists of 4 factors namely rainfall in meters, temperature in °C, pH value of the soil and crop that can be grown. The target will be the yield value of the crop that will be obtained in production / unit area.

TABLE I SAMPLE DATASET OF CROP AND YIELD DATA

1	Rainfall	Temperature	Ph	Crop	Production
2	400.15082	20	3	Bajra	0.0069
3	400.16331	20	3.2	Bajra	0.00747
4	400.16387	20	3.2	Bajra	0.00749
5	400.17967	20	3.2	Bajra	0.00822
6	400.19582	20	3.2	Bajra	0.00895
7	400.29047	20	3.2	Bajra	0.01328
8	400.38869	20	3.2	Bajra	0.01777
9	400.39889	20	3.2	Bajra	0.01824
10	400.41592	20	3.2	Bajra	0.01902

#### B. Essential Packages

- Pandas
- Matplotlib
- Tkinter

- Scikit-Learn

#### C. System Skeleton

System skeleton is portrayed in the figure Fig.1 . below

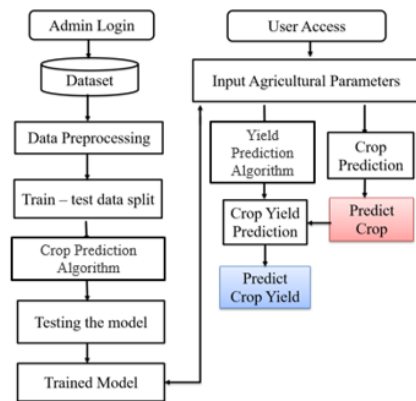


Fig. 1 .System skeleton of the proposed model

#### D. Data Preprocessing

The data present in the dataset is preprocessed to remove redundant values, missing values and other inconsistencies. The target values are also encoded with values between 0 and no\_of\_classes-1 using Label Encoder. This is done for the purpose of classification. Then data is divided into 0.8 of training intended data and 0.2 of testing intended data.

#### E. Crop Prediction

Foremost prediction which the model carries out is that of crop prediction. Crop is predicted based on the historical data of the agricultural fields. The parameters taken into account for prediction are meteorological parameters like rainfall value, temperature value and soil variables like soil pH. In our paper we have done a comparative study of various algorithms that can be used for doing this crop prediction. Below is an account of the algorithms that has been used for prediction purpose.

#### Support vector machines

SVM is basically a learning algorithm that can be used for classification as well as regression purposes. There are different kernel modes that can be used for SVM such as linear, Gaussian and polynomial. But the proposed model is based on SVM with linear kernel since the dataset consists of linearly separable data as shown by the results. The data fits better when linear kernel is chosen.

#### Decision Tree Classifier

Classification and Regression tree is the modern implementation of the decision tree algorithm. The algorithm learns or classifies data on the basis of decision rules which are deduced by training data by calculating entropy and information gain. There will be a tree structure that will be created for classification purpose and each node will represent an attribute. Foremost will be the root node followed by the children nodes. The leaf nodes represent outcome of the decision.

*Random forest Classifier*

It is a supervised learning algorithm that creates numerous instances of decision trees at once based on the observations on the dataset and predicts the output by selecting the decision tree with the most vote. However the disassociation between the different decision trees becomes important. When there is more disassociation it leads to better results.

*k - NN Classifier*

k - NN classifier is a ML algorithm that is formed upon the lazy learning concept. During learning stage the classifications are not generalized in contrast to other machine learning algorithms. The generalizations are only made after users input their prediction queries. The values are predicted on the basis of distance functions.

*F. Prediction of Crop Yield*

The estimation of yield is done using a variety of regression algorithms. Linear Regression received an r2 or Root Mean Square Error score as negative hence was not used. Similarly, the cross validation model ElasticNetCV also was rejected due to a negative score. The Decision Tree Regressor is an algorithm that iteratively learns and builds the decision tree by using the dataset. This algorithm achieved good r2 score when used with the considered dataset compared to other similar regression algorithms. Other notable algorithm was the Random Forest Regressor which achieved similar results and the model was a good fit for the considered dataset.

IV. RESULTS

TABLE II COMPARISON OF VARIOUS ALGORITHMS FOR CROP PREDICTION

Algorithm Used	Crop Prediction Accuracy
Support Vector Machine	92.6%
Decision Tree Classifier	99.87%
k-NN Classifier	99.73%
Random Forest Classifier	81.07%

Table II depicts the various accuracies that is obtained when different algorithms are applied with the crop prediction model.

TABLE III COMPARISON OF VARIOUS ALGORITHMS FOR CROP YIELD PREDICTION

Algorithm Used	Crop Yield Prediction r2 score	Comments
Linear Regression	-6.0358	Worse fit than the horizontal line
Decision Tree Regressor	0.9988	Best fit
ElasticNetCV	-7.2482	Worse fit than the horizontal line
Random Forest Regressor	0.9874	Good fit

Table III depicts the various r2\_scores obtained when different algorithms are applied with the crop yield prediction model.

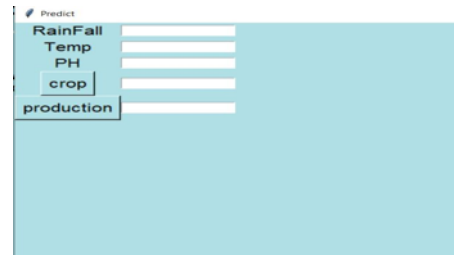


Fig. 2 .User Input

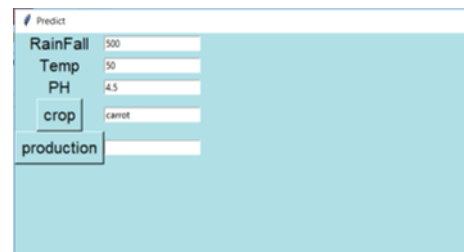


Fig. 3 .Crop Prediction



Fig. 4 .Crop Yield Prediction

V. EPILOGUE AND FORTHCOMING ENHANCEMENTS

This paper demonstrates effective use of various ML algorithms for the crop and crop yield prediction. Prediction of crop and crop yield is done on the basis of input values pertaining to rainfall, soil pH, temperature .Paper also demonstrates comparative study of algorithms with the amount of accuracy and r2 score they obtained and suggesting best algorithm which could be implemented for predicting the crop and yield respectively .The future work is to implement user friendly interface with different languages and also improving the system by using IOT devices.

REFERENCES

[1] Sk Al Zaminur Rahman, Kaushik Chandra Mitra, S.M. Mohidul Islam, "Soil Classification using Machine Learning Methods and Crop Suggestion Based on Soil Series", International Conference of Computer and Information Technology (ICIT), 21-23 December, 2018.  
 [2] Yunous Vagh, Jitian Xiao, "Minimum Temperature Profile Data For Shire-Level Crop Yield Prediction", International Conference on Machine Learning and Cybernetics, Xian, 15-17 July, 2012.

- [3] Igor Oliveria, Renato L, F. Cunha, Bruno Silva, Marco A.S. Netto, "A Scalable Machine Learning System For Pre-Season Agriculture Yield Forecast, IEEE 14<sup>th</sup> International Conference on e-Science, 2018.
- [4] S. Veenadhari, Dr. Bharat Mishra, Dr. CD Singh, "Machine Learning Approach For Forecasting Crop Yield Based On Climatic Parameters", International Conference on Computer Communication and Information, Jan 3-5, 2014.
- [5] Sonal Jain, Dharavath Ramesh, "Machine Learning Covergence For Weather Based Crop Selection", IEEE International Students Conference On Electrical, Electronics and Computer Science, 2020.
- [6] Niketa Gandhi, Leisa J. Armstrong, Owaiz Petkar, Amiya Kumar Tripathy, "Rice Crop Yield Prediction In India Using Support Vector Machines", 13<sup>th</sup> International Joint Conference on Computer Science and Software Engineering (JCSSE), 2016.
- [7] Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Sinng, "Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique", International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials(ICSTM), 6-8 May 2015.
- [8] Grajales D.F.P, Mejia F, Mosquera G.J.A, Piedrahita L.C, Basurto C, "Crop-Planning making smarter agriculture with climate data", "4<sup>th</sup> International Conference On Agro-Geoinformatics", 2015.
- [9] Sahu S, Chawla M, Khare N, "Ancient Analysis Of Crop Yield Prediction Using Hadoop Framework Based On Random Forest Approach", "International Conference On Computing, Communication and Automation(ICCCA), 2017.
- [10] Manjunatha M, Parkavi A, "Estimation Of Arecanut Yield In Various Climatic Zones Of Karnataka Using Data Mining Techniques", International Conference On Current Trends Towards Converging Technologies (ICCTCT).
- [11] BD Parameshchhari et. al "Epileptic Seizure Detection Using Machine Learning," 1<sup>st</sup> International Conference on Emerging Trends in Engineering, Innovative Science and Management (ICETEISM-2019), 2019.