

# Crop Yield Estimation Using Machine Learning Algorithm

**Dr. Shilpa Abhang<sup>1</sup>, Nikhila W<sup>2</sup>**

Assistant Professor, Department of MCA, Jyoti Nivas College, Bangalore, India<sup>1</sup>

Student, Department of MCA, Jyoti Nivas College, Bangalore, India<sup>2</sup>

**Abstract:** Estimation of crop yield former to harvest is a significant concern in agriculture, as the changes in crop yield from year-to-year impact international business, food demand, and global market prices. Also, initial estimation of crop yield renders beneficial report to policy planners. Appropriate estimation of crop productivity is required for proficient planning of land usage and economic policy. The estimation will also help the farmers to make decisions such as the selection of alternative crops or in-order-to discard crops at an initial step in case of crucial situations. Further, estimating crop yield can ease the farmers to have a more-desirable perception on cultivation. Thus, it is essential to simulate and estimate the crop yield before cultivation for effective crop management and expected result. As there exists an affinity between crop yield and the factors influencing crop, machine learning methods may be efficient for yield estimations. In this research work, the Decision tree is applied for the crop yield prediction. The Decision tree is compared with binary SVM classifier and Naïve Bayes. Decision tree performs well for the crop production analysis.

**Keywords:** Decision tree, Binary SVM classifier, Naïve bayes.

## I. INTRODUCTION

Agriculture has a major role in the lives of every individual. From the olden times agriculture is considered to be one of the main practices practiced in India. In olden times, people used to cultivate crops in their own land in order to meet their requirements. It is the broadest commercial sector and plays an important role in comprehensive development of the country. About 60 % of the land in the country is used for agriculture in order to suffice the needs of people. Thus, marketization of agriculture is extremely important and hence will guide the cultivators of our country towards profit. Crop yield estimation in several countries is based on orthodox procedures of data collection for crop and yield estimation depends on field reports. These procedures are expensive, time consuming and are likely to bring-forth huge errors due to deficient surveys, leading to poor crop yield analysis. In utmost countries the data become accessible too late for suitable measures to be taken to avoid food scantiness. Therefore, we need a suitable method or algorithm for crop yield estimation.

The proposed system discusses about estimating yield production. Yield estimation is a very important in agriculture sector for the farmers, prediction will help to overcome situation like drought, rainfall etc. Some of the techniques such as Decision tree. Binary SVM classifier, Naïve bayes can be used to provide the solution for predicting yield production.

## II. LITERATURE REVIEW

In paper [1] authors take into consideration the data related to environmental factors such as by considering various parameter of the soil such as nitrogen, fertility, pH, Phosphate, Potassium and some atmospheric parameter such as sunshine, rainfall and humidity. Above factors hold direct influence on crop yield and also considers past year production and suggests the best profitable crops that can be cultivated in the given environment condition. ANN technique lists out all the possible crops so that the farmer has option to choose the profitable one.

In paper [2] authors perform research by considering parameters such as soil moisture, temperature, humidity and rainfall datasets of all the districts of Telangana State. Support vector machine (SVM), Regression analysis, K Nearest Neighbour (KNN), clustering and various types of techniques are used for prediction. By applying machine learning algorithm, the model has been tested. K-NN suggests suitable accuracy in crop yield prediction. KNN algorithm is used to classify and predict the crop yield.



In paper [3] authors determine the suitable crop to be cultivated and its appropriate yield production. The system gets the input such as state name, district name, season for classifying the crop and inputs such as year and available yield area for predicting the crop production. The model uses decision tree and linear regression algorithms to make predictions of the desired crop and its approximate yield production and Python as the programming language that lets you work quickly and integrate systems more efficiently in machine learning area.

In paper [4] concentrates on estimating the crop yield based on the current data by using Random forest algorithm. Authentic data of Tamilnadu have been utilized for the development of models and the models was examined with samples. Crop yield estimation includes estimating yield of the crop from handy documented data like weather parameters, soil parameters. It achieves an accurate estimation using the Random Forest algorithm. This algorithm consists of a target variable (or dependent variable) which is supposed to be estimated from a given set of predictors (independent variables).

In paper [5] k-means segmentation has been used for yield estimation. It generated a dataset comprising 80 tree photographs with 4000 citrus fruits from different areas. K-means algorithm is popular due to its simplicity and relatively low computational complexity. It is suitable in our circumstances as number of clusters 'K' is simple to select. An photograph of an orange tree generally comprises of regions representing oranges, leaves, branches and sky. Experiments were performed by getting automated counting of citrus using the aimed strategy and comparing the outcomes with the terrain absoluteness. The empirical results show a precise recognition rate of 91.3 %.

### III. PROPOSED METHODOLOGY

This system analyses the Crop yield based on available data in the dataset. The system gets the input such as soil minerals, moisture, temperature (max and min), humidity, rainfall etc. This paper concentrates on estimating the yield of the crop based on the existing data by using Naïve bayes classifier, binary Support Vector machine classifier and Decision tree algorithm.

From experimental results, it has been predicted that Decision tree algorithm found to be most accurate technique for crop yield prediction.

#### Workflow

- Initially load the dataset which is in csv file format using pandas library with all features.
- Clean the data if it has missing values, null values by removing that value in the dataset.
- Train the dataset in the training dataset.
- Apply various algorithms on dataset.
- Predict the results obtained by different algorithms on the training dataset.
- Finally, the algorithms give their respective accuracy of crop yield by considering the features in dataset.

#### Following stages illustrates the step by step approach for the proposed method: -

##### A. Gathering dataset:

Data collection is the method of congregating and measuring details on variables of interest. Data is congregated from various resources and it is stored.

##### B. Acquisition of training dataset:

The accuracy of a machine learning algorithm may depend on several of parameters used and to the extent of correctness of the dataset. Our dataset contains the soil minerals, moisture, temperature (max and min), humidity, rainfall etc. Thus, by using an appropriate machine learning algorithm we can train the dataset to predict the most suitable crop that can be grown under the given input parameters.

##### C. Data pre-processing:

Data pre-processing is a process that is used to transform the rough data into clean data. Data pre-processing is the second step and it contains two steps. Original dataset can contain lots of missing values so initially all these should be removed. Missing values are denoted by a dot in the dataset and their presence can deteriorate the value of entire data and it can reduce the performance. So, to solve this problem we replace these values with large negative values which will be treated



as outliers by the model. Generating the class labels is the second step. Since we are using a supervised learning method, for each entry in the dataset there should be a class label which is created during the pre-processing step.

#### D. Models/algorithms selection for comparison:

We need to examine, collate and select the good one that suits this particular dataset. Generally, when solving a machine learning problem with a specified dataset, then test various models and methods to interpret an optimization problem and fit the highly suitable model.

#### E. Machine Learning Algorithm:

Machine Learning is a significant decision support tool for crop yield estimation. This algorithm is the measurement that is used to determine which model is good at detecting relations and patterns between parameters in dataset based on the input data. Different Machine Learning algorithms are then used to induce comparisons.

## IV. IMPLEMENTATION

### Environmental Setup

This research is implemented in Python SciPy environment installed ideally with Python 3.8.3 and Jupyter lab and notebook is utilized for execution. Libraries like Scikit-learn, Pandas, NumPy is installed. Minimum of 8GB RAM is required.

### Experimental Results

From experimental results, it can be found out that all Machine learning techniques under study can be used for crop yield prediction. Decision tree attend to be the method of selection for predictive patterning because they are comparatively simple to understand and also very impactful. The primary objective of a Decision tree is to divide a population of data into tiny partitions Decision tree algorithm predicted highest accuracy of 99.4% whereas Naïve Bayes classifier and Binary SVM classifier predicted average accuracy of 82% and 89% respectively.

#### i.Naïve Bayes

---

```
Confusion Matrix:
[[ 9  1  0  4]
 [ 0  5  0  0]
 [ 0  0 11  0]
 [ 0  0  2  7]]
Accuracy: 0.8205128205128205%
```

---

#### i.Binary SVM classifier

```
Confusion Matrix:
[[870  6  1  0  72]
 [ 41 346  2  0 114]
 [ 23  2 549  0  0]
 [  3  1  0  26  6]
 [ 47  8  0  0 865]]
Accuracy: 0.8906773977196513%
```

---

#### i.Decision Tree

```
Accuracy: 0.9949748743718593
```



## VI. CONCLUSION

In this proposed system, we have applied Decision tree Algorithm and predicted the crop yield in which it is getting 99.4% of accuracy by taking statistical raw datasets which contains various fields like Average Humidity, Mean Temp, Min Temp, Moisture, Alkaline, Chalky, sandy, Max Temp, Millet Yield, Rainfall and, Sandy. As python is easy which has run time results and Jupyter is useful for the entire implementation is done by using Anaconda Software which consists of both Python IDLE and Jupyter Notebook.

## REFERENCES

1. Karan Chaudhary, et al, *Prediction of crop yield using machine learning*,2020, Vol. 4, Issue 9, International Journal of Engineering Applied Sciences and Technology
2. S. Pavani, Augusta Sophy Beulet P, et al, *Heuristic Prediction of Crop Yield using Machine Learning Technique*,2019, Volume-9 Issue-1S3, International Journal of Engineering and Advanced Technology
3. Aksheya Suresh, K. Monisha, R. Pavithra,B. Marish Hariswamy, et al, *Crop Selection and it's Yield Prediction*,2020, Volume-8 Issue-6, International Journal of Recent Technology and Engineering
4. Niketa Gandhi, Owaiz Petkar, et al, *CS771 PredictingRice Crop Yield Using Bayesian Networks*,2016, Conference on Advances in Computing, Communications and Informatics (ICACCI)
5. Miss.Snehal S.Dahikar, Dr.Sandeep V.Rode, et al, *Agricultural Crop Yield Prediction Using Artificial Neural Network Approach*,2014, Vol. 2, Issue 1.
6. Kusum Lata, et al, *crop yield prediction using data mining techniques and machine learning models for decision support system*,2019, Volume 6, Issue 4.
7. Vaishali Pandith, et al, *Performance Evaluation of Machine Learning Techniques for Mustard Crop Yield Prediction from Soil Analysis*, 2020, Volume 64, Issue 2, 2020 Journal of Scientific Research.
8. P.Priya, et al, *predicting yield of the crop using machine learning algorithm*,2018, International Journal Of Engineering Sciences & Research Technology.
9. Zeeshan Malik, Sheikh Ziauddin, Ahmad R. Shahid, and Asad Safi, *Detection and Counting of On-Tree Citrus Fruit for Crop Yield Estimation*,2016, Vol. 7, No. 5, 2016, (IJACSA) International Journal of Advanced Computer Science and Applications.
10. Vinita Shah, Prachi Shah, et al, *Groundnut Crop Yield Prediction Using Machine Learning Techniques*, 2018, Volume 3 , Issue 5, International Journal of Scientific Research in Computer Science, Engineering and Information Technology.
11. Prithvi Ram, Rajeshwari S.B, Jagadish S Kalliman, et al, *Smart Yield Accuracy Prediction using Linear Regression and Collaborative Filtration*,2019, Volume-8 Issue-3, International Journal of Recent Technology and Engineering (IJRTE).
12. Saeed Khaki and Lizhi Wang, *Crop Yield Prediction Using Deep Neural Networks*, 2019, Industrial and Manufacturing Systems Engineering, Iowa State University, Ames, IA, United States.