

# Prediction of Crop Yield and Cost by Finding Best Accuracy by Using Machine Learning Approach

**Mrs.K.Amsavalli<sup>1</sup>, V.Sundhar<sup>2</sup>, B.Suresh<sup>3</sup>**

Assistant Professor, CSE, Anand Institute of Higher Technology, Chennai, India<sup>1</sup>

Student, B.E., Computer Science and Engineering, Anand Institute of Higher Technology, Chennai, India<sup>2,3</sup>

**Abstract:** Among worldwide, agriculture has the major responsibility for improving the economic contribution of the nation. However, still the most agricultural fields are under developed due to the lack of deployment of ecosystem control technologies. Due to these problems, the crop production is not improved which affects the agriculture economy. Hence a development of agricultural productivity is enhanced based on the plant yield prediction. To prevent this problem, Agricultural sectors have to predict the crop from given dataset using machine learning techniques. The analysis of dataset by Supervised Machine Learning Technique (SMLT) to capture several information's like, variable identification, uni-variate analysis, bi-variate and multi-variate analysis, missing value treatments etc. A comparative study between machine learning algorithms had been carried out in order to determine which algorithm is the most accurate in predicting the best crop. The results show that the effectiveness of the proposed machine learning algorithm technique can be compared with best accuracy with entropy calculation, precision, Recall, F1 Score, Sensitivity, Specificity and Entropy.

**Keywords:** Machine learning, Entropy calculation, Precision, Dataset.

## I. INTRODUCTION

In developing countries, farming is considered as the major source of revenue for many people. In modern years, the agricultural growth is engaged by several innovations, environments, techniques and civilizations. In addition, the utilization of information technology may change the condition of decision making and thus farmers may yield the best way. For decision making process, data mining techniques related to the agriculture are used. Data mining is a process of extracting the most significant and useful information from the huge number of datasets. Nowadays, we used machine learning approach with developed in crop or plant yield prediction since agriculture has different data like soil data, crop data, and weather data. Plant growth prediction is proposed for monitoring the plant yield effectively through the machine learning techniques. It is also applicable for the automated process of farming is the beginning of a new era in Bangladesh that will be suitable for the farmers who seek experts to take suggestion about the appropriate crop on specific location of their land and don't want to forget any step of the cultivation throughout the process. Although, the opinion from experts is the most convenient way, this application is designed to give accurate solution in fastest manner possible. This research's main objective is to bring farming process a step closer to the digital platform.

## II. EXISTING SYSTEM

It presents a crop/weeds classification approach based on a three-steps procedure. The first step is a robust pixel-wise segmentation (i.e., soil/plant) and image patches containing plants are extracted in the second step. The third step, a deep CNN for crop/weed classification is used. The extracted blobs in the masked image containing plants information are fed to a CNN classifier based on a fine-tuned model of VGG-16 exploiting the ability of deep CNN in object classification and to reduce the limitations of CNNs in generalizing when a limited amount of data is available. The classification step can then be specialized to the types of plants needed by the application scenario. It evaluated the complete pipeline, including the first background removal phase and the subsequent classification stage. Experimental results demonstrate that can achieve good classification results on challenging data. Precision agriculture is gaining increasing attention because of the possible reduction of agricultural inputs (e.g., fertilizers and pesticides) that can be obtained by using high-tech equipment, including robots. To focus on an agricultural robotics system that addresses the weeding problem by means of selective spraying or mechanical removal of the detected weeds. To describe a deep learning based method to allow a robot to perform an accurate weed/crop classification using a sequence of two Convolutional Neural Networks (CNNs) applied to RGB images. The first network, based on encoder-decoder segmentation architecture, performs a pixel wise, plant-type agnostic, segmentation between vegetation and soil that enables to extract a set of connected blobs representing plant instances.

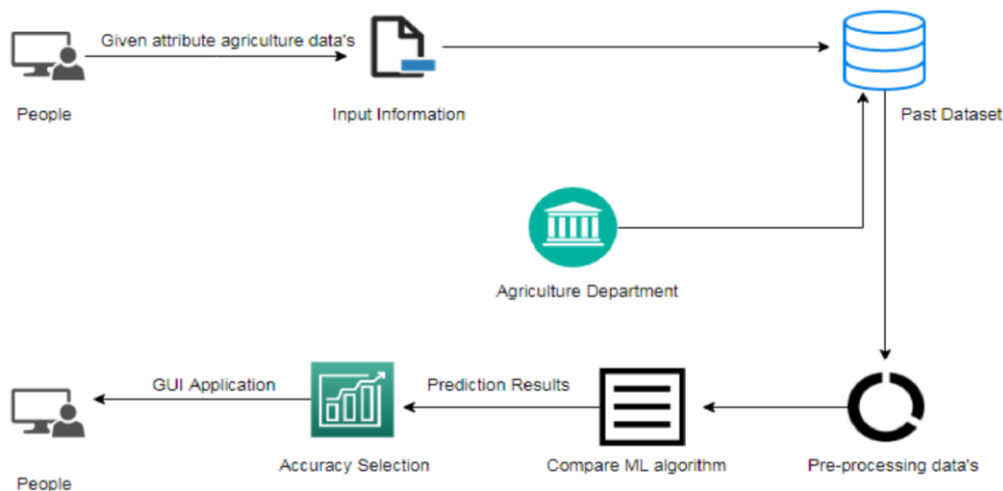
### III. PROPOSED SYSTEM

In this section of the report, you will load in the data, check for cleanliness, and then trim and clean your dataset for analysis. Make sure that you document your steps carefully and justify your cleaning decisions.

- The first line imports iris data set which is already predefined in sklearn module. Iris data set is basically a table which contains information about various varieties of iris flowers.
- For exmp, to import any algorithm and train\_test\_split class from sklearn and numpy module for use in this program.
- Then we encapsulate load\_data() method in data\_dataset variable. Further we divide the dataset into training data and test data using train\_test\_split method. The X prefix in variable denotes the feature values and y prefix denotes target values.
- This method divides dataset into training and test data randomly in ratio of 67:33. Then we encapsulate any algorithm.
- In the next line, we fit our training data into this algorithm so that computer can get trained using this data. Now the training part is complete.

### IV. SYSTEM ARCHITECTURE

This is the most exciting phase in Applying Machine Learning to any Dataset. It is also known as Algorithm selection for Predicting the best results. Usually Data Scientists use different kinds of Machine Learning algorithms to the large data sets. But, at high level all those different algorithms can be classified in two groups: supervised learning and unsupervised learning. Supervised learning: Supervised learning is a type of system in which both input and desired output data are provided. Input and output data are labelled for classification to provide a learning basis for future data processing. Supervised learning problems can be further grouped into **Regression** and **Classification** problems.



### V. SYSTEM IMPLEMENTATION

#### Module-01: Variable Identification Process / data validation process:

Validation techniques in machine learning are used to get the error rate of the Machine Learning (ML) model, which can be considered as close to the true error rate of the dataset. If the data volume is large enough to be representative of the population, you may not need the validation techniques. However, in real-world scenarios, to work with samples of data that may not be a true representative of the population of given dataset. To finding the missing value, duplicate value and description of data type whether it is float variable or integer.

#### Module-02: Exploration data analysis of visualization:

Data visualization is an important skill in applied statistics and machine learning. Statistics does indeed focus on quantitative descriptions and estimations of data. Data visualization provides an important suite of tools for gaining a qualitative understanding. This can be helpful when exploring and getting to know a dataset and can help with identifying patterns, corrupt data, outliers, and much more. With a little domain knowledge, data visualizations can be used to express and demonstrate key relationships in plots and charts that are more visceral and stakeholders than measures of association or significance. Data visualization and exploratory data analysis are whole fields themselves and it will recommend a deeper dive into some the books mentioned at the end.

**Module-03: Logistic Regression:**

It is a statistical method for analysing a data set in which there are one or more independent variables that determine an outcome. The outcome is measured with a dichotomous variable (in which there are only two possible outcomes). The goal of logistic regression is to find the best fitting model to describe the relationship between the dichotomous characteristic of interest (dependent variable = response or outcome variable) and a set of independent (predictor or explanatory) variables. Logistic regression is a Machine Learning classification algorithm that is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data coded as 1 (yes, success, etc.) or 0 (no, failure, etc.).

**Module-04: Support Vector Machines (SVM):**

A classifier that categorizes the data set by setting an optimal hyper plane between data. I chose this classifier as it is incredibly versatile in the number of different kernelling functions that can be applied and this model can yield a high predictability rate. Support Vector Machines are perhaps one of the most popular and talked about machine learning algorithms. They were extremely popular around the time they were developed in the 1990s and continue to be the go-to method for a high-performing algorithm with little tuning.

**Module-05: K-Nearest Neighbor**

K-Nearest Neighbor is a supervised machine learning algorithm which stores all instances correspond to training data points in n-dimensional space. When an unknown discrete data is received, it analyzes the closest k number of instances saved (nearest neighbors) and returns the most common class as the prediction and for real-valued data it returns the mean of k nearest neighbors. In the distance-weighted nearest neighbor algorithm, it weights the contribution of each of the k neighbors according to their distance using the following query giving greater weight to the closest neighbors.

**VI. APPLICATIONS**

- It is an integrated farm management application using mobile app.
- Agricultural sector to automate to identify the crop prediction process (real time world) and predicting by desktop application / web application.

**VII. CONCLUSION**

The analytical process started from data cleaning and processing, missing value, exploratory analysis and finally model building and evaluation. Finally, we predict the crop using machine learning algorithm with different results. This brings some of the following insights about crop prediction. As maximum types of crops will be covered under this system, farmer may get to know about the crop which may never have been cultivated and lists out all possible crops, it helps the farmer in decision making of which crop to cultivate. Also, this system takes into consideration the past production of data which will help the farmer get insight into the demand and the cost of various crops in market.

**VIII. FUTURE ENANCEMENT**

- Remaining SMLT algorithm will be involve to finding the best accuracy with applying to predict the crop yield & cost.
- Agricultural department wants to automate the detecting the yield crops from eligibility process (real time).
- To automate this process by show the prediction result in web application or desktop application.
- To optimize the work to implement in Artificial Intelligence environment.

**REFERENCES**

- [1]. Sa,Z.Chen,M.Popovi´c,R.Khanna,F.Liebisch,J.Nieto,andR.Siegiwart, “weednet: Dense semantic weed classification using multispectral images and mav for smart farming,” IEEE Robotics and Automation Letters, vol. 3, no. 1, pp. 588–595, 2018.
- [2]. M. Di Cicco, C. Potena, G. Grisetti, and A. Pretto, “Automatic model based dataset generation for fast and accurate crop and weeds detection,” in IROS, 2017, pp. 5188–5195.
- [3]. E. Shelhamer, J. Long, and T. Darrell, “Fully convolutional networks for semantic segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 4, pp. 640–651, 2017.
- [4]. P. Lottes, J. Behley, N. Chebrolov, A. Milioto, and C. Stachniss, “Joint stem detection and crop-weed classification for plant-specific treatment in precision farming,” arXiv preprint arXiv:1806.03413, 2018.
- [5]. C. McCool, T. Perez, and B. Upcroft, “Mixtures of lightweight deep convolutional neural networks: Applied to agricultural robotics,” IEEE Robotics and Automation Letters, vol. 2, no. 3, pp. 1344–1351, 2017.
- [6]. W. Strothmann, A. Ruckelshausen, J. Hertzberg, C. Scholz, and F. Langsenkamp, “Plant classification with in-field-labeling for crop/weed discrimination using spectral features and 3d surface features from a multi-wavelength laser line profile system,” Computers and Electronics in Agriculture, vol. 134, pp. 79–93, 2017.
- [7]. S. Haug and J. Ostermann, “A crop/weed field image dataset for the evaluation of computer vision based precision agriculture tasks,” in Computer Vision - ECCV 2014 Workshops, 2015, pp. 105–116.