Prediction of Crop Yield And Crop Cost By Finding The Best Accuracy

S.Akshaya[1], G.Kaviya[2], U.Krithisha[3] ,R.Aparna[4], R.K. Kapilavani[5]
[1, 2, 3] Dept of Computer Science and Engineering
[4, 5] Assistant Professor, Dept of Computer Science and Engineering
[1, 2, 3, 4, 5] Prince Shri Venkateshwara Padmavathy Engineering College

Abstract- *This Paper proposes a methodology, which aims to predict the crop yield and crop cost by given dataset using machine learning technique. In worldwide, the major responsibility for improving the economic contribution of the nation mainly depends upon the agriculture. Due to the lack of deployment of ecosystem control technologies, still most of the agricultural fields are under developed .Because of these problems, production of crop is not improved which affects the agriculture economy. The plant yield prediction enhanced the development of agricultural productivity. The given dataset is analyzed by supervised machine learning technique (SMLT) to get the information like variable identification, uni-variate analysis, bi-variate and multi-variate analysis, missing values and null values etc. The best crop is predicted by the comparative study among the machine learning algorithms. The algorithm with best accuracy is chosen with the entropy calculation, precision, Recall, F1 Score, Sensitivity, Specificity and Entropy*

Keywords- Dataset, machine learning –classification method.

I. INTRODUCTION

Faming is considered as the major source of the revenue for many people in the developing countries .In modern days the growth of the agriculture is improved by several innovations, techniques and civilizations. Decision making of the farmers may change by the use of information technology in agriculture. Data mining technique related to agriculture is used for the decision making process. Data mining is the process of getting the most significant and useful information from the huge dataset. Now a days we use machine learning techniques to predict the crop yield .Because to predict the crop we need the several data like soil data, crop data, and weather data.

Machine learning is the technique which is used to predict the future from past data. It is a type of Artificial intelligence (AI). Machine learning techniques provide the computer the ability to learn without being explicitly programmed. Machine learning is categorized into three types. There are supervised learning, unsupervised learning and reinforcement learning. In supervised learning both input data and labeling is given .In unsupervised learning no label is given. It figures out clustering of the input data. Reinforcement learning receives positive and negative feedback to improve its performance. This application is designed to give accurate solution in fastest manner. This research‘s main objective is to bring farming process a step closer to the digital platform

II. RELATED WORK

This is section we discuss about literature survey, existing system, proposed system and architecture framework.

Machine Learning is the process of training the computer with the past data set and predicts the output. The process of training and prediction involves use of specialized algorithms. It feed the training data to an algorithm, and the algorithm uses this training data to give predictions on a new test data.

Supervised machine learning algorithm is used. The algorithm is Random forest, Decision tree, Logistic Regression, Support vector machine.

III. LITERATURE SURVEY

A. Automated Farming Prediction

Various machine learning algorithm is compared to determine the best algorithm which give most accurate in predicting the best crop in particular land [1].The best crop is the crop which has most increase in terms of yield per unit area. The algorithms used are Multiple Linear Regression (MLR) and K-Nearest Neighbor Regression (KNNR).Multiple Linear Regression produces accurate results.

Crop prediction and cultivation are not digital. There is no collaboration between agriculture and technology.

B. Machine Learning Approaches to Corn Yield Estimation Using Satellite Images and Climate Data

Various Machine learning techniques such as Support vector machine, Random forest, Extremely Randomized Trees and Deep Learning is used. To determine the corn yield estimation. Machine Learning is an efficient empirical method for classification and prediction.[2] This paper we examine seasonal sensitivity of corn fields. Temporal characteristics of crop yield are not analyzed. Tests to examine climate changes affecting crop yield are not made.

C. Prediction of Crop yield using Big data

Quantifying the yield to ensure food security using big data analysis [3]. Algorithm involves map-reduce data processing structure and Nearest Neighbors for modeling. In Nearest Neighbors results gained from the former data processing structure, provides a well balanced result on the account of accuracy and prediction time in advance. Comparing the weather similarity and data processing in controlled computational time.

D. Rice Crop Yield Prediction in India using Support Vector Machines

Food production in India is dependent on various cereal crops such as rice, wheat and various pulses. Sustainability and productivity are dependent on climatic conditions in order to make decisions. variation in seasonal climate conditions can have detrimental effect[4].Developing better techniques to predict crop productivity in different climate conditions can help the farmer to take the better decision in crop choice. Support Vector Machine technique is used. Other classifiers such as Naïve Bayes, Bayes Net provide better accuracy and quality than Sequential Minimal Optimization (SMO).

E. Use of Data Mining in Crop Yield Prediction

Data mining technique is used to analyze the agricultural dataset [5]. Classifier such J48, Locally Weighted Learning, Logical Analysis of Data (LAD),Instance Based K-nearest neighbor are used. Performance of each classifier are compared using WEKA tool by various parameters such as Root Mean Squared Error(RMSE),Mean Absolute Error(MAE) and Relative Absolute Error(RAE).Data Mining can handle only small datasets.

IV. EXISTING SYSTEM

Classification of crop/weeds method is based on three steps In First step segmentation of robust pixel-wise. In Second step plants which contain image patches are extracted. In third step Deep CNN is used for crop weed classification. 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. Types of plants needed by application are specialized in the classification step. 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.

The agriculture input such as fertilizers and pesticides have been reduced in precision agriculture by using high-technology equipment such as robots. Agricultural robots remove the weeding problem by means of selective spraying or mechanical removal of detected weeds. Accurate weed/crop is classified by robot by using two Convolution Neural Network (CNN) which is applied to RGB images. Pixel wise, plant type agnostic, segmentation between vegetation and soil is performed by encoder-decoder segmentation architecture.

DRAWBACKS

1. 1 .It can't determine to improve the classification accuracy of our pipeline.
2. 2. Connecting the bridge manually and some corruption are happened.
3. 3. Private sectors domination high, profit low and credits not getting concern farmer.

V. PROPOSED SYSTEM

Four different types of machine learning algorithm are used. Logistic Regression, Random Forest, Support Vector Machine, Decision Tree. Among the four algorithms .Algorithm with best accuracy is chosen .Best accuracy algorithm provides the farmer and idea to decide which crop have to be too planted. Past data set of crop details is given has the input to the computer to predict the output. Past data set contains states name, district name, crop year, season, crop, area, rainfall, average humidity, mean temperature, cost of cultivation, cost of production, yield (Quintal, hectare) and cost of production per yield.

1. Logistic Regression

Logistic Regression is multiple type of classification algorithm. It can be probability score, underlying the model classification.

It is used to predict the probability of categories of dependent variable. Dependent variable is binary variable coded as 0 or1.

Data collected is grouped by crop details and our goal is to predict type crop, crop yield and cost production.

2. Decision Tree

Decision tree is a supervised machine learning algorithm. It is most powerful and popular algorithm. Trained dataset crop details are considered as root node. Attributes are assumed to be categorical for information gain and for gini index. Decision tree also work well in continuous values. Attributes are ordered in ascending order.

3. Random Forest

Random Forest is a supervised machine learning algorithm based on ensemble learning. It combines multiple algorithm of same type. It can be used of classification and regression tasks.

4. Support Vector Machine

Support Vector Machine is the most popular machine learning algorithm. To separate multiple classes of data, many possible hyper planes can be chosen. Best plane can be selected based on maximum margin.

7 parameters are common in Logistic regression, Random forest, and Support vector machine and decision tree.

Parameters are precision, recall, f1-score, support, sensitivity, specificity, confusion matrix.

1. Formula for calculating precision

$$\text{Precision} = \frac{TP}{TP+FP}$$

2. Formula for calculating Recall or sensitivity

$$\text{Recall} = \frac{TP}{TP+FN}$$

5. Formula for calculating f1-score

$$\text{F1-score} = \frac{Precision*Recall}{2*Precision+Recall}$$

6. Formula for calculating specificity

$$\text{Specificity} = \frac{TN}{FP+TN}$$

5. Formula for calculating confusion matrix

TP	FN
FP	TN

Ensemble Algorithm

Ensemble algorithm is the combination of machine learning algorithm. It combines algorithms of same type multiple times or different type of algorithms to build a more powerful prediction model.

Ensemble algorithm is used to improve the machine accuracy.

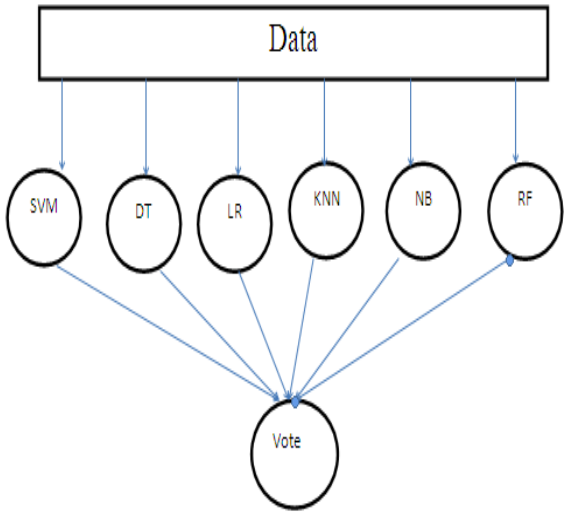
There are 3 types

1. Stacking
2. Bagging
3. Boosting

We had used stacking method. Stacking method is based on voting.

Max voting

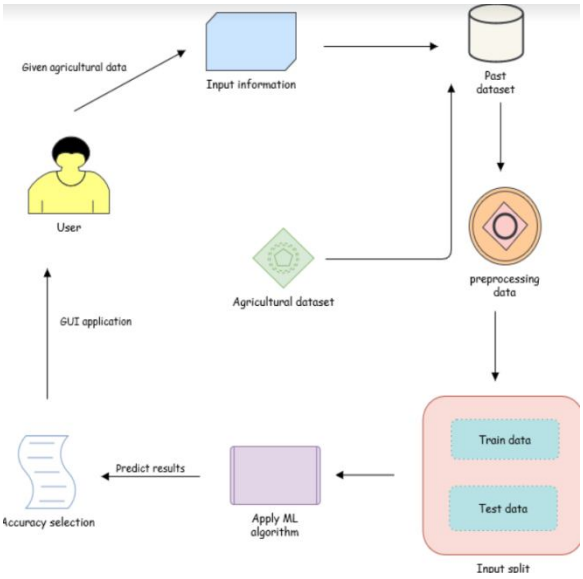
The max voting method is generally used for classification problems. Multiple models are used to make predictions for each data point. The predictions by each model are considered as vote. The predictions which we get from the majority of the models are used as the final predictions.



Ensemble Structure

VI. SYSTEM ARCHITECTURE

The system has the main participants: the user and the agricultural department. The user will define the problem to the system that provides the GUI for the user. The required features are extracted from this problem to create the input information. This information is compared with the past agricultural dataset. Input data is preprocessed by splitting of input information into train data and test data. The four Machine Learning algorithms: Decision Tree, Random Forest, Logistic Regression and Support Vector Machine are used to predict crop yield and cost. This process is followed by accuracy prediction. Here, various parameters are used to classify the best Machine Learning algorithm that produces result with best accuracy is chosen as the final output using Ensemble Learning.



VII. MODULES

1) Data validation and preprocessing technique.
2) Exploration data analysis of visualization and training a model by given attributes.
3) Performance measurements of ML algorithm based on crop yield.
4) Performance measurements of ML algorithm based on crop cost.
5) Performance measurements of Ensemble algorithm.
6) A) GUI based prediction of ranking of crops.
   B) GUI based prediction of crop yield and yield cost.

MODULE 01

Data validation and preprocessing technique

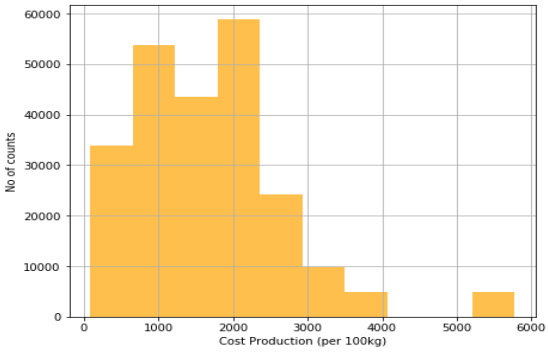
The Validation techniques in machine learning are used to get the error rate of the Machine Learning (ML) model, which can be considered close to the true error rate of the dataset. If the data volume is large enough to be representative of the population, you might not need the validation techniques. Pre-processing refers to the transformations applied to our data before feeding it to the algorithm. Data preprocessing, here, is a technique that is used to convert the given raw data into a clean data set. Or, whenever the data is gathered from different sources it is collected in raw format which is not feasible for analysis. To achieve better results from the applied model in Machine Learning method the data has to be in a proper manner. Some specified Machine Learning models need information in a specified format; for example, Random Forest algorithm does not allow null values. Hence, to execute random forest algorithm, null values have to be reduced from the original raw data set.

MODULE 02

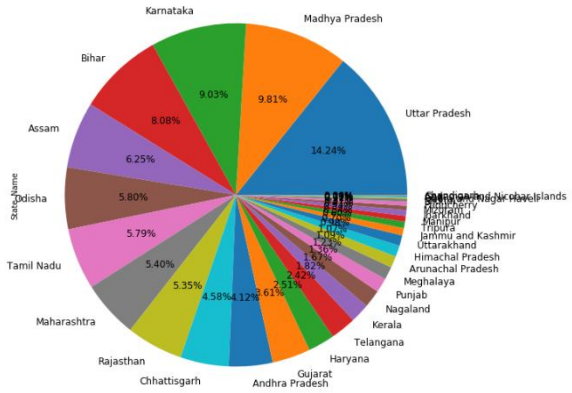
Exploration data analysis of visualization by training a model by given attributes

Data visualization is an important process in applied statistics and machine learning. Sometimes data will not make sense until it can be looked at in a visual form, such as with charts and plots.

Cost production per 1000kg by counts



Percentage of crop yield production by state



MODULE 03 & 04

Performance measurements of ML algorithm based on crop yield

Logistic Regression

It is a statistical method in which there are one or more independent variables that determine an outcome. The goal of logistic regression is to find the best model to describe the relationship between the dependent variables and a set of independent variables.

Decision Tree

Decision tree creates classification or regression models in the form of a tree structure. It breaks a data set into smaller subsets while at the same time an associated decision tree is developed. Decision tree builds classification or regression models in a tree structure.

Support Vector Machine

Support Vector Machines are one of the most popular and talked about machine learning algorithms. I chose this method 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.

Random Forest

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees.

MODULE 05

Ensemble learning

Ensemble learning helps improve machine learning results by combining several models. This approach allows the production of better predictive performance compared to a single model and it is the art of combining diverse set of learners (individual models) together to improvise on the stability and predictive power of the model.

Max Voting: The max voting method is used for classification problems. Here, multiple models are used to make predictions for each data point. The predictions by each model are considered as 'votes'. The predictions which we get from the majority of the models are used as the final predicted value.

Averaging: Here, multiple predictions are made for each data point in averaging. We take the average of predictions from all the models and use it to make the final prediction.

MODULE-06

GUI for crop yield and cost prediction

Tkinter is the python library for developing GUI (Graphical User Interfaces). We use this library for creating an application of UI (User Interface), to create windows and all other graphical user interface and Tkinter comes with Python as a standard package, it can be used for security purpose for each user or accountant. There will be two kinds of pages like registration user purpose and login entry purpose of users.

VIII. CONCLUSION

The analytical process started from data cleaning and processing, exploratory analysis and finally model building and evaluation. Finally, we predicted the crop using machine learning algorithms with different results. This brings some insights about crop prediction. As maximum crops will be covered under this system, a 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. This system also takes into consideration the past production of data which will help the farmer to get insights about the demand and the cost of various crops in market. To improve the above system, we can implement it in an Artificial Intelligence environment to perform using real-time data.

REFERENCES

[1] Snehal Chaudhary, Shruti Mishra, Priyanka Paygude, Sonali Idate "Use of Data Mining in Crop Yield

Prediction“ In Proceedings of the Second International Conference on Inventive Systems and Control, 2018.

[2] Dipro Barua, Talha Siddique, Zannatul Ferdous, Amitabha Chakrabarty, “Automated Farming Prediction” In Intelligent Systems Conference, 2017.

[3] Leisa J. Armstrong, Niketa Gandhi, Owaiz Petkar, Amiya Kumar Tripathy “Rice Crop Yield Prediction in India using Support Vector Machines” In 13th International Joint Conference on Computer Science and Software Engineering , 2016.

[4] Nari Lee, Kim, Yang-Won “Machine Learning Approaches to Corn Yield Estimation Using Satellite Images and Climate Data: A Case of Iowa State” In Journal of the Korean Society of Surveying, Geodesy, Photogrammetry and Cartography, 2016.

[5] Wu Fan , Chen Chong , Guo Xiaoling , Yu Hua, Wang Juyun “Prediction of crop yield using big data” In 8th International Symposium on Computational Intelligence and Design, 2015.