Prediction of Agriculture Crop Yield and Product Cost by GUI Application using Random Forest Algorithm

NAVANEETHA KIRSHNAN M, Head of the Department,

Department of Computer Science and Engineering

KUMARAVEL M, Student of Computer Science and Engineering

St. Joseph College of Engineering, Sriperumbudur, Chennai.

Abstract:

Among worldwide, agriculture has the major responsibility for improving the economic contribution of the nation. However, still the most agricultural fields are underdeveloped 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 must 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 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 – Crop recommendation, pre-processing, classifier algorithm, feature extraction, Machine Learning.

Introduction:

In order to perform predictive analysis, Agricultural systems are particularly efficient when it comes to increasing productivity and crop production efficiency. The population, on the other hand, grows slowly, while crop production resources dwindle day by day.

Traditionally, farming entails either planting or harvesting a crop according to a set schedule.

Real-time meteorological data, air quality, soil, crop maturity, machinery, labour costs, and current data availability are all required for precision agriculture. In today's digital environment, we use a more intelligent decision-making process to solve our problems. As a result, in the agricultural sector, we apply a prediction model.

Farmers forecast crop production or yield based on their experience; nevertheless, this is not the best technique. Consultant oculus observation is the most widely used approach for plant or crop forecast and identification.

Objectives:

- The reason for this downturn in the agriculture industry is that farmers are not being paid properly. It's because of the absence of application of IT in the agriculture sector.
- The goal is to develop a machine learning model for Crop yield Prediction, to potentially replace the updatable supervised machine learning classification models by predicting results in the form of best accuracy by comparing supervised algorithm.

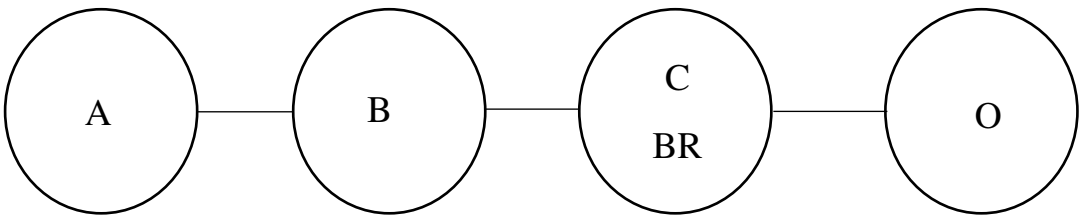
Literature Survey:

Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. **Author:** Douglas K. Bolton∗, Mark A. Friedl. We developed empirical models forecasting maize and soybean production in the Central United States using data from NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) in combination with county-level data from the United States Department of Agriculture (USDA). We also investigated MODIS' ability to capture inter-annual variations in yields as part of our research.

Crop Yield Assessment from Remote Sensing.**Author:** Paul C. Doraiswamy, Sophie Moulin, Paul W. Cook, and Alan Stern.The US Department of Agriculture is very interested in crop condition and production estimates at the state and county levels. The United States Department of Agriculture's National Agricultural Statistical Service (NASS) conducts field interviews with sampled farm operators and collects crop cuttings to estimate crop yields at the regional and state levels.

"Machine learning approach for forecasting crop yield based on climatic parameters " **Author:** S.Veenadhari, Dr. Bharat Misra, Dr. CD Singh .In the sphere of agriculture, climate plays a significant effect. The climate has been severely impacted this year as a result of increased global warming, and this has had a significant influence on agriculture. According to the predictive analysis, Predicting crop production will tell farmers what to harvest.

MATHEMATICAL MODELING:



A = User entered input,

B = Pre-process,
C = Feature selection,
BR = Pre-process request evaluation,
O = Predict outcome.

Set Theory

1) Let S be as system which input image S = {In, P, Op, }
2) Identify Input In as
In = {A}

Where,
A = User entered input (dataset)

3) Identify Process P as P = {B, C, BR}

Where,
B = Pre-process
C = Feature selection
BR = Pre-process request evaluation

4) Identify Output Op as
Op = {UB}

Where,
UB = Predict outcome =Failures and Success conditions.

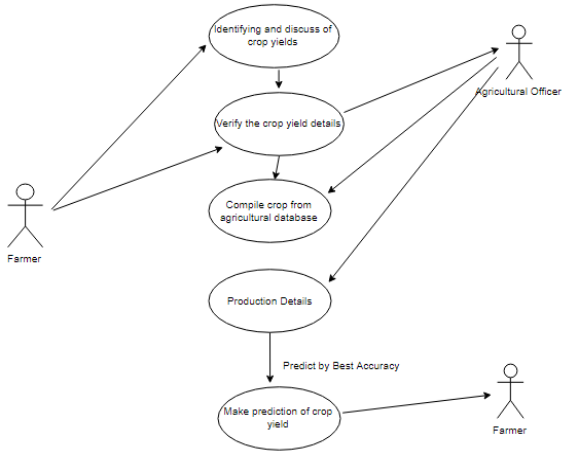
System Design:

Exploratory Data Analysis:

I will load in the data, check for cleanliness, and then trim and clean the dataset for analysis in this phase of the report. Make sure you meticulously record your steps and provide justifications for your cleaning decisions.

Training the Dataset:

For example, to use in this programme, import any algorithm and train test split class from the sklearn and numpy modules. The load data() method is thus encapsulated in the data dataset variable. Using the train test split function, we further divide the dataset into training and test data. The feature values are denoted by the X prefix in variable, while the target values are denoted by the y prefix. This method randomly divides the dataset into training and test data at a 67:33 ratio. Then any algorithm is encapsulated.

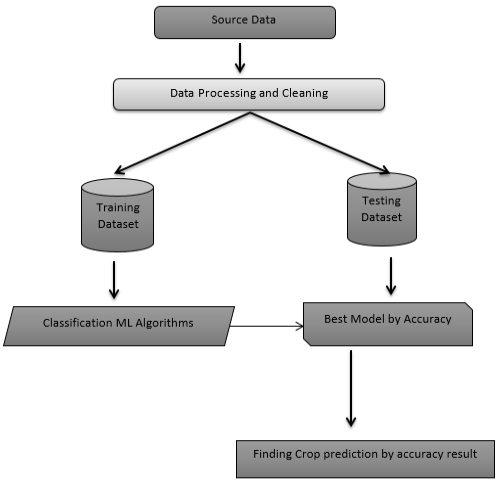


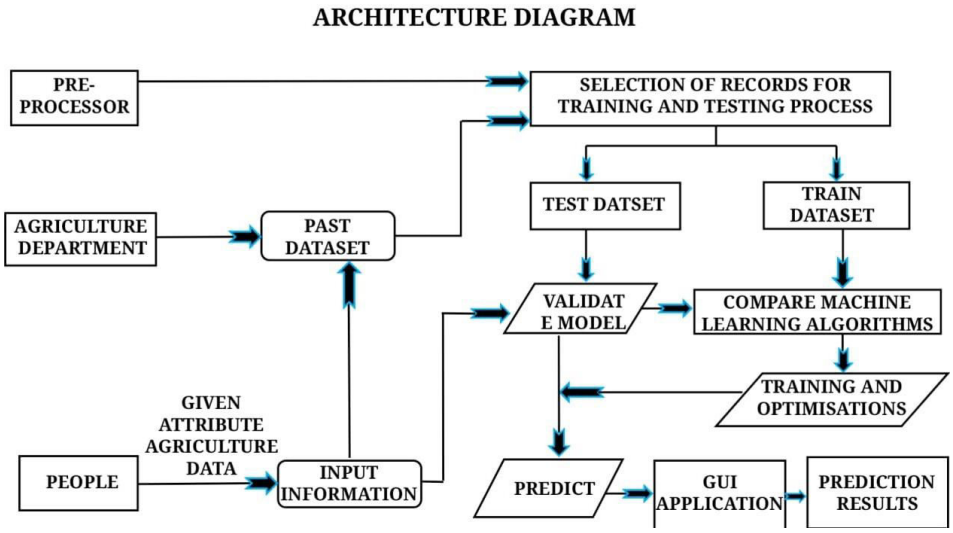
Testing the Dataset

• We now have the measurements of a new blossom in a numpy array called 'n,' and we want to guess what species it is. This is accomplished by utilising the predict method, which accepts this array as an input and returns the predicted target value as an output.

• As a result, the expected target value is 0. Finally, we calculate the test score, which is the ratio of correct predictions to total guesses. We do this using the scoring technique, which compares thetest set's actual values to the projected values.

Work Flow Diagram





So the complete work flow can able to get the maximum accuracy level of values for the prediction project.

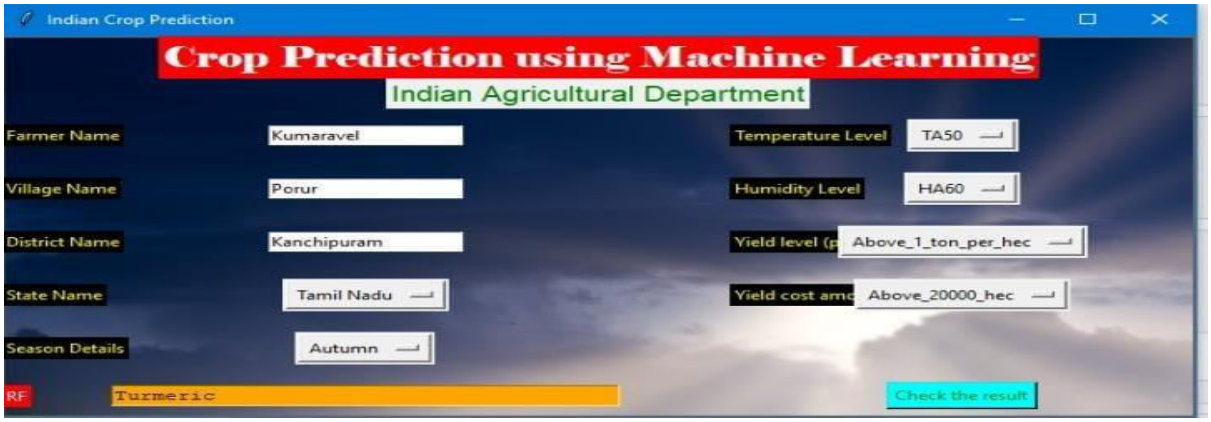
Implementation:

Algorithm Explanation:

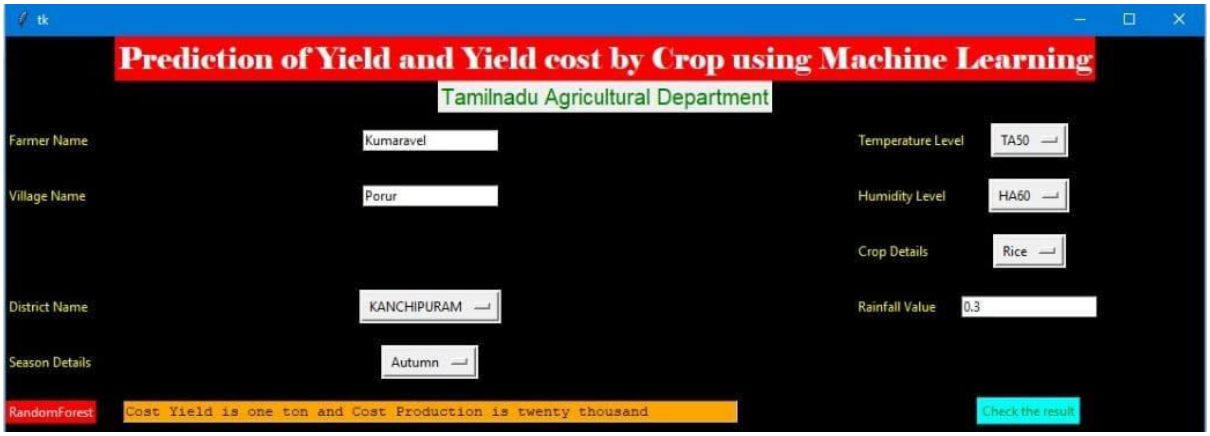
```
# Importing the required libraries
import pandas as pd,
numpy as np import matplotlib.pyplot as plt, seaborn as sns
%matplotlib inline
# Reading the csv file and putting it into 'df' object
df = pd.read_csv('heart_v2.csv')
df.head()
# Putting feature variable to X
X = df.drop('heart disease',axis=1)
# Putting response variable to y
y = df['heart disease']
# now lets split the data into train and test
from sklearn.model_selection import train_test_split
# Splitting the data into train and test
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size=0.7,
random_state=42) X_train.shape, X_test.shape
from sklearn.ensemble import RandomForestClassifierclassifier
rf=RandomForestClassifier(random_state=42,n_jobs=-1,max_depth=5,
n_estimators=100, oob_score=True)
%%time classifier_rf.fit(X_train, y_train)
imp_df = pd.DataFrame
({ "Varname": X_train.columns,
"Imp": rf_best.feature_importances_
})
```

```
imp_df.sort_values(by="Imp", ascending=False)
```

The output for the Crop prediction is must be a clear cut one for the assumption of the crop which is getting from the place.



The output for the Product cost prediction is must be a clear cut one for the assumption of the crop which is getting high cost from the place.



Conclusion:

Data cleaning and processing, missing value analysis, exploratory analysis, and model creation and evaluation were all part of the analytical process. Finally, we use a machine learning method to predict the crop, with varying outcomes. This leads to some of the following crop forecast insights. Because this system will cover the most sorts of crops, farmers will be able to learn about crops that have never been farmed before and will be able to see a list of all possible crops, which will aid them in deciding which crop to cultivate. Furthermore, this method takes into account previous data production, allowing the farmer to gain insight into market demand and costs for particular crops.

Future Enhancement:

- The remaining SMLT algorithms will be used to determine the optimum accuracy for predicting agricultural yield and cost.
- The agricultural department seeks to automate the process of recognising high-yield crops (real time).
- This method can be automated by displaying the prediction result in a web or desktop application.
- To optimise the labour for Artificial Intelligence implementation.

Dr.M.Navaneethakrishnan M.E., PhD is a Head of the Department in the Department of Computer Science and Engineering at St. Joseph College of Engineering, Sriperumbudur, Chennai, Tamil Nadu. He has completed his Ph.D, in Cyber Security - Computer Science and Engineering in 2017 from Manonmaniam Sundaranar University (MSU) Tirunelveli, Tamilnadu. He has done his M.E, CSE in Anna University Chennai in the year 2008. Dr.M.Navaneethakrishnan has 15 years of teaching experience and has 58 publications in International Journals and Conferences. His research interests include network security, Computer Networks, data science and Machine Learning. He is an active member of ISTE, CSI, IEANG and IEI

Mr.M.Kumaravel B.E.,Student of Computer Science and Engineering at St.Joseph College of Engineering,Sriperumbudur,Chennai,TamilNadu.I had attended many International Conference, Workshops, Hackathons and Seminars in the area of Data Science Organizations,Python ,Machine Learning And Deep learning Respectively. I got Placed in Reputed Companies like Infosys, Capgemini and some respeted companies.