PREDICTION OF CROP YIELD AND COST ESTIMATION DETECTION USING MACHINE LEARNING

TEAM MEMBERS:

MUKIL CHOKALINGAM M (211417104154)
NAVEEN PR (211417104164)
NAVEEN NARAYAN M (211417104162)

PROJECT GUIDE:

Mr. KAJENDRAN K
M.C.A., M.E.,

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.

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.
- Machine learning method is a process of extracting the most significant and useful information from the huge amount 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.

LITERATURE SURVEY

Title	Objectives	Techniques used	Merits
Deep Learning Classification for Crop Types in North Dakota (2020)	Predict crop for particular land	Deep Neural Network, Crop mapping, Image Classification	Cropland Data Layer (CDL) highly accurate on major crop types (e.g., ~95% accuracy for corn and soybean)
Sorghum Yield Prediction using Machine Learning (2019)	Prediction of yield using Machine Learning	Neural Networks and Linear Regression	RNN
Estimating Crop Yield from Multi-temporal Satellite Data Using Multivariate Regression and Neural Network Techniques (2019)	Development of objective mathematical models of crop yield prediction	Neural Networks Techniques	Three statistical parameters are used for performance analysis: correlation coefficient (*r*), root mean square error

Title	Objective	Techniques used	Merits
A Granular GA-SVM Predictor for Big Data in Agricultural Cyber-Physical Systems(2019)	To achieve result using Granular GA-SVM	SVM -Support Vector Machine	The granulation technique is used to overcome the low efficiency of SVM for large-scale data
Use of Data Mining in Crop Yield Prediction(2018)	crop yield prediction system by using Data Mining techniques	Data Mining	The system to use LAD Tree
Crop Predicting Using Predictive Analytics(2017)	Predicting crop using Predictive analytics	predictive modeling, machine learning and data mining	1. Various soil samples taken from different placed can be tested. 2.Portable

PROBLEM STATEMENT

- Agriculture is the most important sector that influences the economy of India. It contributes to 18% of India's Gross Domestic Product (GDP) and gives employment to 50% of the population of India.
- To focuses on implementing crop yield prediction system by using Machine learning techniques by doing analysis on agriculture dataset.
- For evaluating performance Accuracy is used as one of the factors.
- The classifiers are further compared with the values of Precision, Recall and F1score.
- Lesser the value of error, more accurate the algorithm will work.
- The result is based on comparison among the classifiers.

Existing System

- 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.

Drawbacks:

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

Proposed System

- We have to find Accuracy of the training dataset, Accuracy of the testing dataset, Specification, False Positive rate, precision and recall by comparing algorithm using python code. The following Involvement steps are,
- Define a problem
- Preparing data
- Evaluating algorithms
- Improving results
- Predicting results

DEVELOPMENT ENVIRONMENT(H/W and S/W)

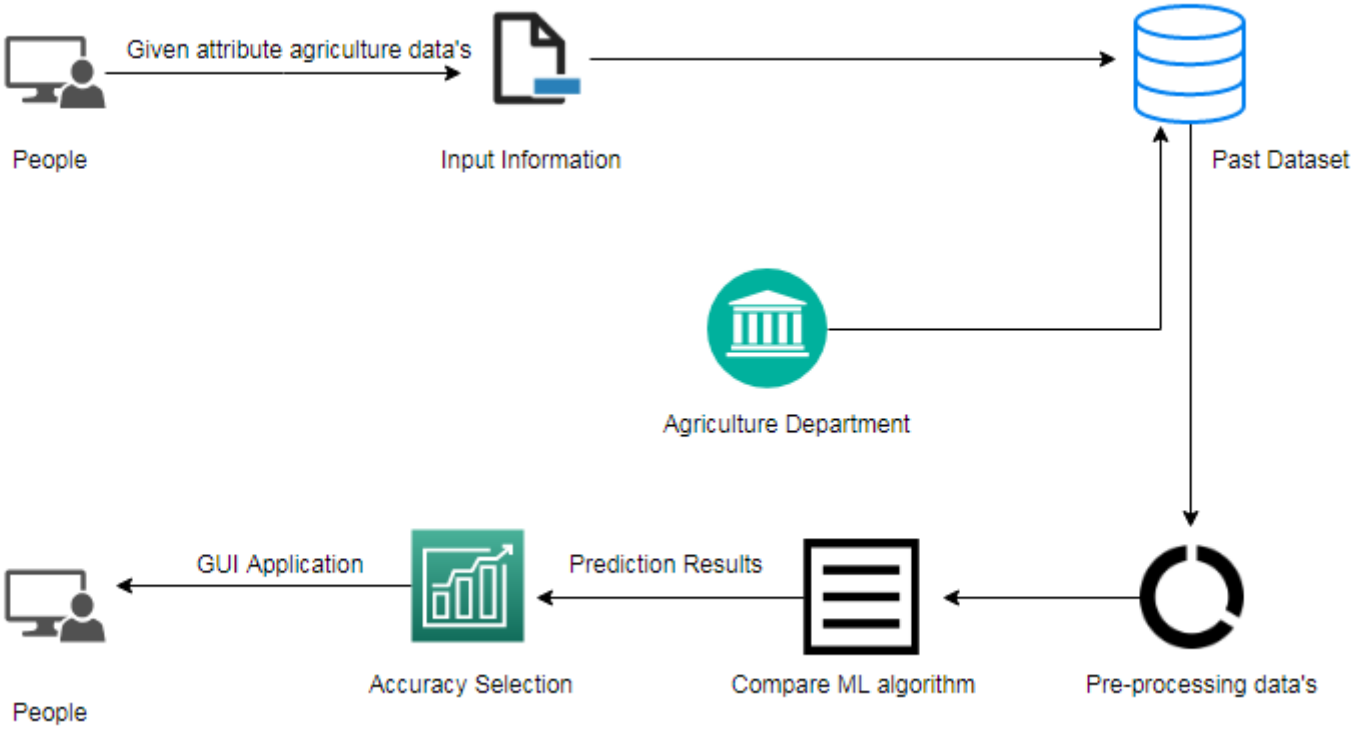
HARDWARE REQUIREMENTS

- Processor – i3
- RAM - 4 GB
- Hard Disk - 300 GB

SOFTWARE REQUIREMENTS

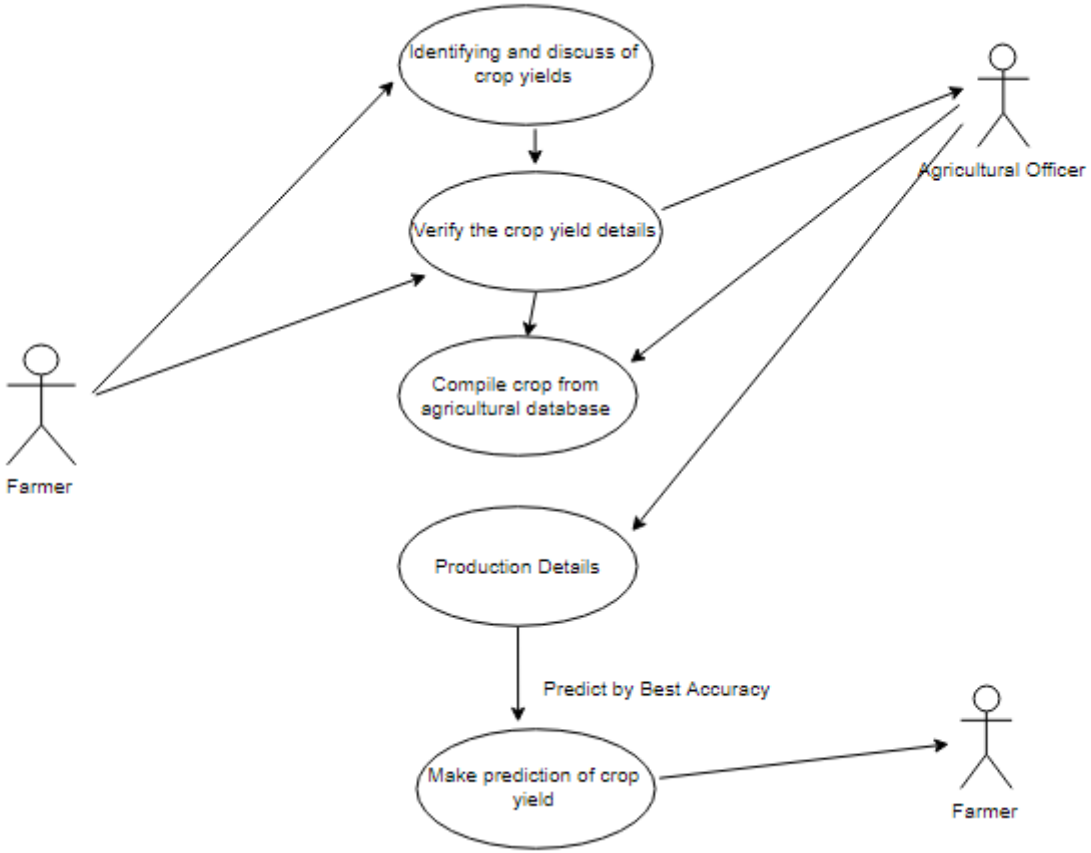
- Operating System - Windows /LINUX
- Tool –Anaconda with Jupiter Notebook

SYSTEM ARCHITECTURE



SYSTEM DESIGN

USE CASE DIAGRAM



MODULE DESCRIPTION

- Data validation and pre-processing technique (Module-01)
- Exploration data analysis of visualization and training a model by given attributes (Module-02)
- Performance measurements of logistic regression and decision tree algorithms (Module-03)
- Performance measurements of Support vector classifier and Random forest (Module-04)
- Performance measurements of KNN and Naive Bayes (Module-05)
- GUI based prediction of crop yield and yield cost (Module-06)

Data Validation and Preprocessing Techniques

- 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.

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.

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

Support Vector Machines

- 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

K-Nearest Neighbour

- *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.

GUI based prediction of crop yield and yield cost

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

TESTING

- Testing is a process of executing a program with the intent of finding an error. A successful test is one that uncovers an as-yet- undiscovered error. System testing is the stage of implementation, which is aimed at ensuring that the system works accurately and efficiently as expected before live operation commences. It verifies that the whole set of programs hang together.
- The software testing process commences once the program is created and the documentation and related data structures are designed. Otherwise the program or the project is not said to be complete. A good test case design is one that as a probability of finding an yet undiscovered error. A successful test is one that uncovers an yet undiscovered error. Any engineering product can be tested in one of the two ways:
- Black Box testing
- White Box testing

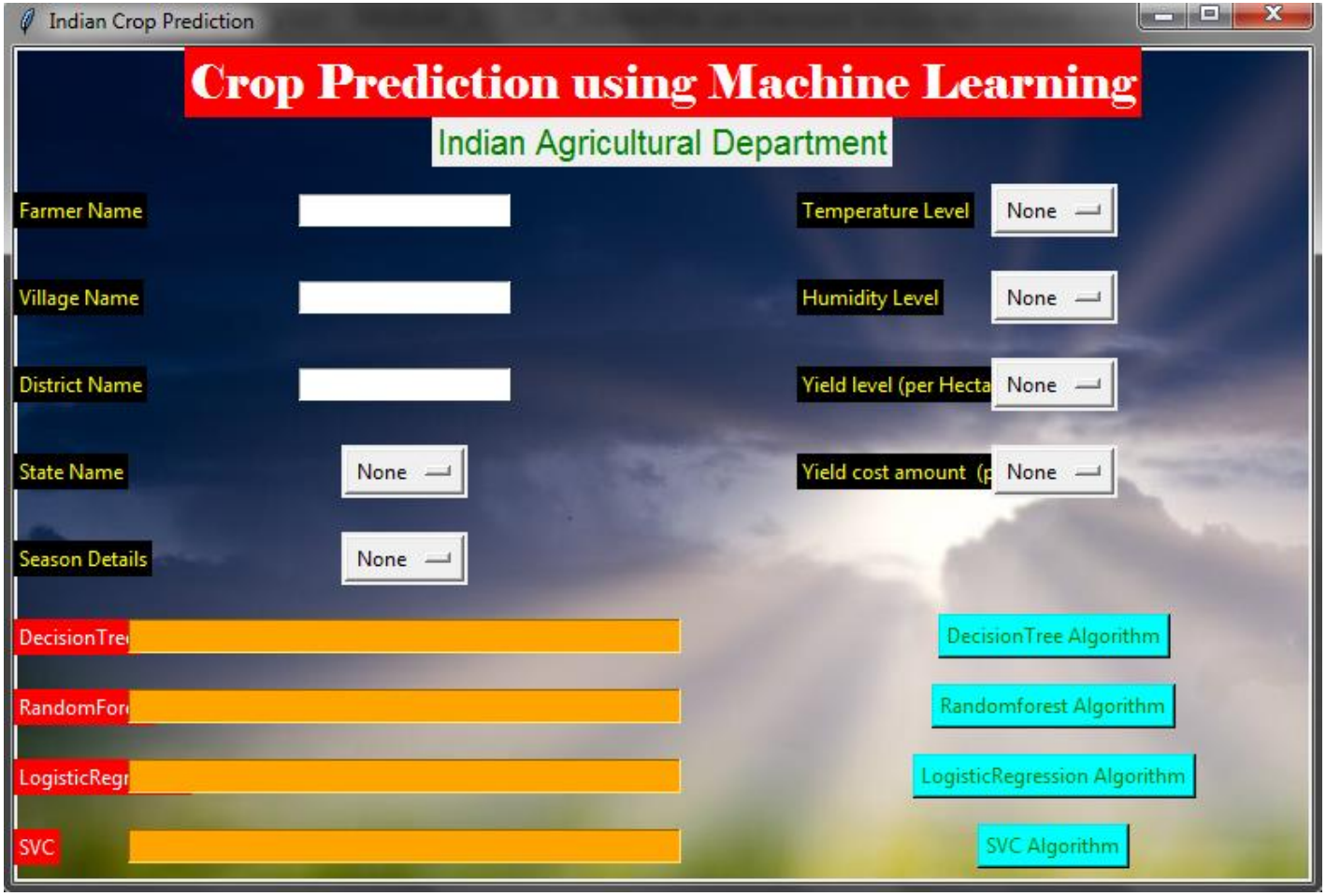
APPLICATIONS

- 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

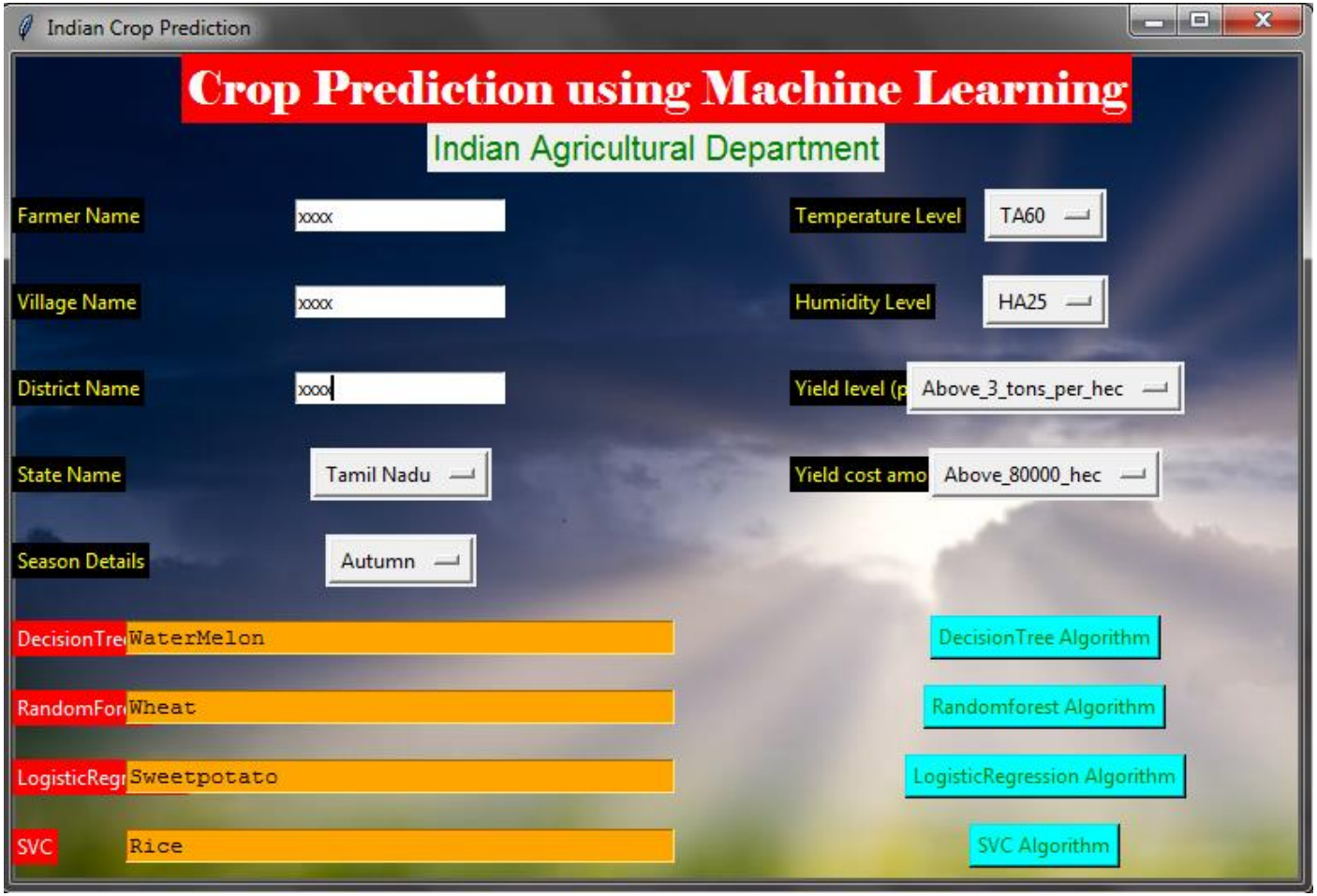
CONCLUIONS AND FUTURE ENHANCEMENTS

- Agricultural department wants to automate the detecting the yield crops from eligibility process (real time).
- To optimize the work to implement in Artificial Intelligence environment.

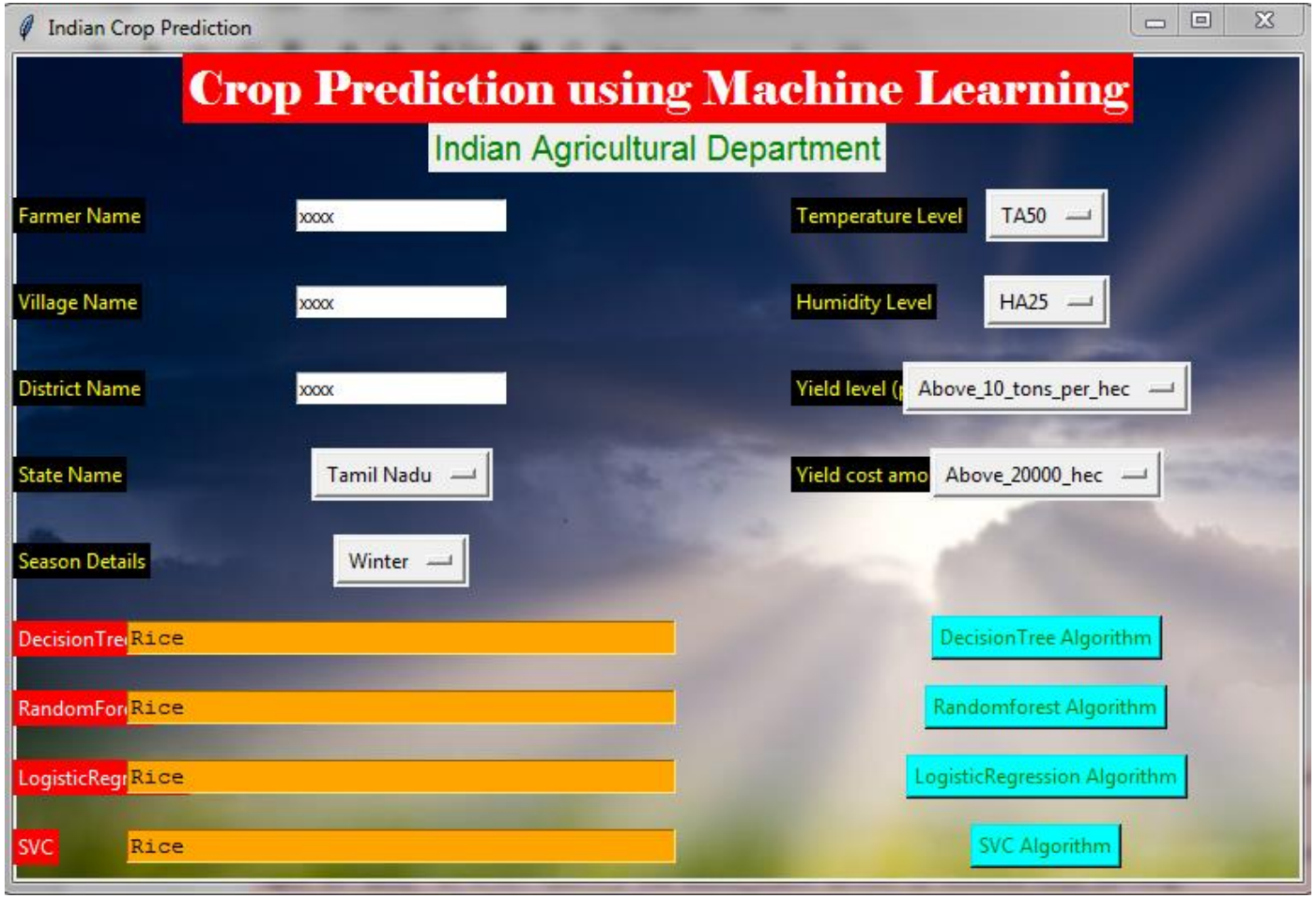
Testing Results :
Input:



Output:
Test 01:



Test 02:


Indian Crop Prediction
Crop Prediction using Machine Learning
Indian Agricultural Department
Farmer Name
xxxx
Village Name
xxxx
District Name
xxxx
State Name
Tamil Nadu
Season Details
Winter
Temperature Level
TA50
Humidity Level
HA25
Yield level (
Above_10_tons_per_hec
Yield cost amo
Above_20000_hec
DecisionTre
Rice
RandomFor
Rice
LogisticRegr
Rice
SVC
Rice
DecisionTree Algorithm
Randomforest Algorithm
LogisticRegression Algorithm
SVC Algorithm

REFERENCES

- D. J. Mulla, "Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps," *Biosyst. Eng.*, vol. 114, pp. 358–371, 2013.
- S. K. Seelan, S. Laguette, G. M. Casady, and G. A. Seielstad, "Remote sensing applications for precision agriculture: A learning community approach," *Remote Sens. Environ.*, vol. 88, pp. 157–169, 2003.
- P. J. Pinter Jr. *et al.*, "Remote sensing for crop management," *Photogram- metric Eng. Remote Sens.*, vol. 69, pp. 647–664, 2003.
- L. Di, G. Y. Eugene, L. Kang, R. Shrestha, and Y.-Q. Bai, "RF-CLASS: A remote-sensing-based flood crop loss assessment cyber-service system for supporting crop statistics and insurance decision-making," *J. Integrative Agriculture*, vol. 16, pp. 408–423, 2017.
- M. A. Friedl *et al.*, "Global land cover mapping from MODIS: algorithms and early results," *Remote Sen. Environ.*, vol. 83, pp. 287–302, 2002.

References

- X. Zhang *et al.*, “Monitoring vegetation phenology using MODIS,” *Remote Sen. Environ.*, vol. 84, pp. 471–475, 2003.
- F. Gao *et al.*, “Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery,” *Remote Sen. Environ.*, vol. 188, pp. 9–25, 2017.
- C. Boryan, Z. Yang, R. Mueller, and M. Craig, “Monitoring US agri-culture: the US department of agriculture, national agricultural statistics service, cropland data layer program,” *Geocarto Int.*, vol. 26, pp. 341–358, 2011.

PAPER PUBLICATIONS

- Journal Name-International Journal of Advanced Research in Computer And Communication Engineering(IJARCCE)
- Paper ID:DOI: 10.17148/IJARCCE.2021.105135
- Published in : Volume 10 Issue 5

THANK YOU!