# Project 1: Prediction of Agriculture Crop Production In India

SIVAGNANASEKAR G 1 , SHAMEEM MOHAMED A R 2

[99210041286@klu.ac.in](mailto:99210041286@klu.ac.in) 1  , [sham2000md@gmail.com](mailto:sham2000md@gmail.com) 2

Prediction of agriculture production in INDIA India produces a wide variety of crops, which are essential to the nation's economy and food security. One of the biggest growers of numerous crops worldwide is India. A sizable majority of the nation's workforce is employed in agriculture, which also accounts for a sizable portion of the GDP.

Artificial Intelligence :

Artificial intelligence is the term used to describe the creation of intelligent machines that can carry out tasks that traditionally require human intelligence. Machine Learning :

Predict the future with the help of existing data.

There are lots of techniques used to predict the crop products , let us see one by one : Regression analysis :

Regression is mostly used for predict the future output. Based on the data we can use regression . Many types of regressions are avilable like simple linear , multiple linear regression etc. Here we use the dependent and independent data to predict the future output.

Decision Trees:

Hierarchical models called decision trees can be used to forecast agricultural yields. Decision trees can create predictions by taking several paths through the tree by building a tree-like structure from input data. Decision trees are helpful for understanding the important elements impacting crop yield since they can handle both numerical and categorical data and are interpretable.

Random Forests: To increase prediction accuracy, random forests is an ensemble learning technique that blends different decision trees. For addressing intricate interactions between input variables and crop yield, random forests are effective. In comparison to individual decision trees, they can handle big datasets, provide feature importance rankings, and lessen overfitting. Support Vector Machines (SVM):

SVM is a potent machine learning method used for regression and classification tasks. SVM can be used to forecast crop yield and determine the correlation between input factors and crop yield. SVM seeks to identify an ideal hyperplane that divides classes or forecasts continuous values. Nonlinear relationships can be handled via kernel functions.

Neural networks:

Deep learning models in particular have acquired favour in a number of disciplines, including the prediction of crop yield. In order to assess crop health and forecast yield, convolutional neural networks (CNNs) can scan satellite or drone data. Recurrent neural networks (RNNs) have the ability to simulate sequential data and detect temporal trends in environmental or meteorological variables.

What we choose to solve our model :

We choose random forest method to predict our future prediction and we use python GUI to display the output .

WEEK 2

Random Forest :

Popular machine learning technique Random Forest is utilised for both classification and regression tasks. It is a method of ensemble learning that mixes different decision trees to generate predictions. Here are some crucial details regarding the Random Forest approach:

Random Forest is an illustration of ensemble learning, in which various models are integrated to enhance overall performance. Decision trees serve as the individual models in Random Forest.

Multiple decision trees are constructed by Random Forest based on various subsets of the training data. A random subset of the characteristics and samples from the original dataset is used to train each tree individually.

Random Forest introduces randomization in two different ways. At each node of the decision tree, a subset of features is first chosen at random. The training data is then randomly sampled with replacement using bootstrapping. These methods of randomization aid in reducing overfitting and strengthening the model.

Bagging: Random Forest builds a variety of decision trees using a method known as bagging (bootstrap aggregating). By randomly choosing samples with replacement, bagging entails extracting numerous bootstrap samples from the original dataset. One of these bootstrap samples is used to train each tree.

After all the decision trees have been trained, the forecasts are combined by Random Forest to get the final prediction. It employs majority voting for classification problems, selecting the class that has received the most votes from the individual trees. It uses the average of the projected values from each individual tree for regression tasks.

Random Forest is able to quantify the significance of a feature. It determines the typical impurity reduction brought on by a characteristic across all of the decision trees, such as the Gini impurity or entropy. This data can be helpful for feature selection and determining the relative weights of various dataset features.

Overfitting-resistant: Random Forest is renowned for being overfitting-resistant. Combining different decision trees using randomization methods reduces variation and enhances generalisation performance. In comparison to individual decision trees, it is less susceptible to noisy data and outliers.

Scalability: Random Forest is capable of processing big datasets with multidimensional attributes. Individual decision trees can be trained in parallel, making it possible to implement them quickly on multicore processors and distributed computing frameworks.

The number of trees in the forest, the maximum depth of each tree, the number of features taken into account at each split, and other hyperparameters can all be tweaked to improve performance in Random Forest. To prevent either underfitting or overfitting, careful hyperparameter optimisation is essential.

Interpretability: Although Random Forest offers metrics of feature relevance, the total model is more difficult to understand than a single decision tree. Because the predictions are based on the collective judgements of several trees, it might be difficult to pinpoint the precise logic behind a given forecast.

Due to its accuracy, resilience, and adaptability, Random Forest is a strong and popular machine learning technique. It can be used in many different fields, including as finance, healthcare, image analysis, and natural language processing.

GUI (tkinter)

The common Python package Tkinter is used to build GUIs, or graphical user interfaces. It offers a collection of widgets and features that enable programmers to create interactive applications with a unified look and feel. The following are some significant Tkinter points:

Integration: The Tcl programming language gave rise to the Tk GUI toolkit, which is the foundation of Tkinter. Since it is part of Python, using Tkinter does not require the installation of any other packages or dependencies.

Cross-platform: Tkinter is a Python GUI development tool that runs on a variety of operating platforms, including Windows, macOS, and Linux.

User interfaces can be made using a variety of built-in widgets that Tkinter offers. These comprise buttons, labels, text boxes, check boxes, radio buttons, entry fields, and more. The use of frames, grids, or pack layouts can be used to order and organise widgets.

Event-driven programming: Tkinter uses this approach to programming. It enables programmers to associate particular user actions, such as button presses or keyboard input, with particular functions or processes. This makes it possible to develop applications that are interactive and responsive.

Geometry management: Tkinter provides three separate techniques—pack, grid, and place—for controlling the positioning and organisation of widgets within a window or frame. Widgets are automatically arranged horizontally or vertically via the pack approach. Widgets can be arranged using the grid approach in a grid-like structure with rows and columns. Widget size and position can be precisely controlled with the put technique.

Customization: Tkinter offers tools for altering the way widgets look. When creating an application, developers can change variables like size, colour, font, and style to achieve the desired appearance and feel. Widgets can also show images, providing further customisation opportunities.

Event loop: To handle events and maintain the responsiveness of the GUI, Tkinter uses an event loop. The accompanying event handlers or callbacks are triggered by the event loop, which monitors user activity. It makes that the programme is always running and capable of responding to user input.

Support for dialogues and message boxes: Tkinter has built-in support for dialogues and message boxes that are often used, including file dialogues, colour pickers, alert boxes, and boxes that request user confirmation. These pre-made dialogues offer a practical way to communicate with users and get information.

Extensibility: Tkinter can be made more functional and integrated with other programmes and frameworks by integrating new Python libraries and modules. It can be used, for instance, in conjunction with Matplotlib to develop interactive data visualisation for Tkinter applications.

Resources and documentation: Tkinter comes with a wealth of information, including official Python documentation and several online tutorials, guidelines, and examples. Developers in need of assistance or motivation can find resources and support from the vibrant Tkinter community.

The popular and adaptable Tkinter package is used to build GUI programmes in Python. It offers a simple and approachable method for creating user interfaces, making it suited for both inexperienced and seasoned developers.

WEEK 3

Popular machine learning algorithm Random Forest is a part of the supervised learning methodology. It can be applied to ML issues involving both classification and regression. It is built on the idea of ensemble learning, which is a method of integrating various classifiers to address difficult issues and enhance model performance.

According to what its name implies, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead than depending on a single decision tree, the random forest uses forecasts from each tree and predicts the result based on the votes of the majority of predictions.

Assumptions of random forest :

* There should be some actual values in the feature variable of the dataset so that the classifier can predict accurate results rather than a guessed result.
* The predictions from each tree must have very low correlations.

The Working process :

Step-1: Select random K data points from the training set.

Step-2: Build the decision trees associated with the selected data points (Subsets).

Step-3: Choose the number N for decision trees that you want to build.

Step-4: Repeat Step 1 & 2.

Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.

Prediction

Training

Data set 2

Training

Data set 1

Training

Data set n

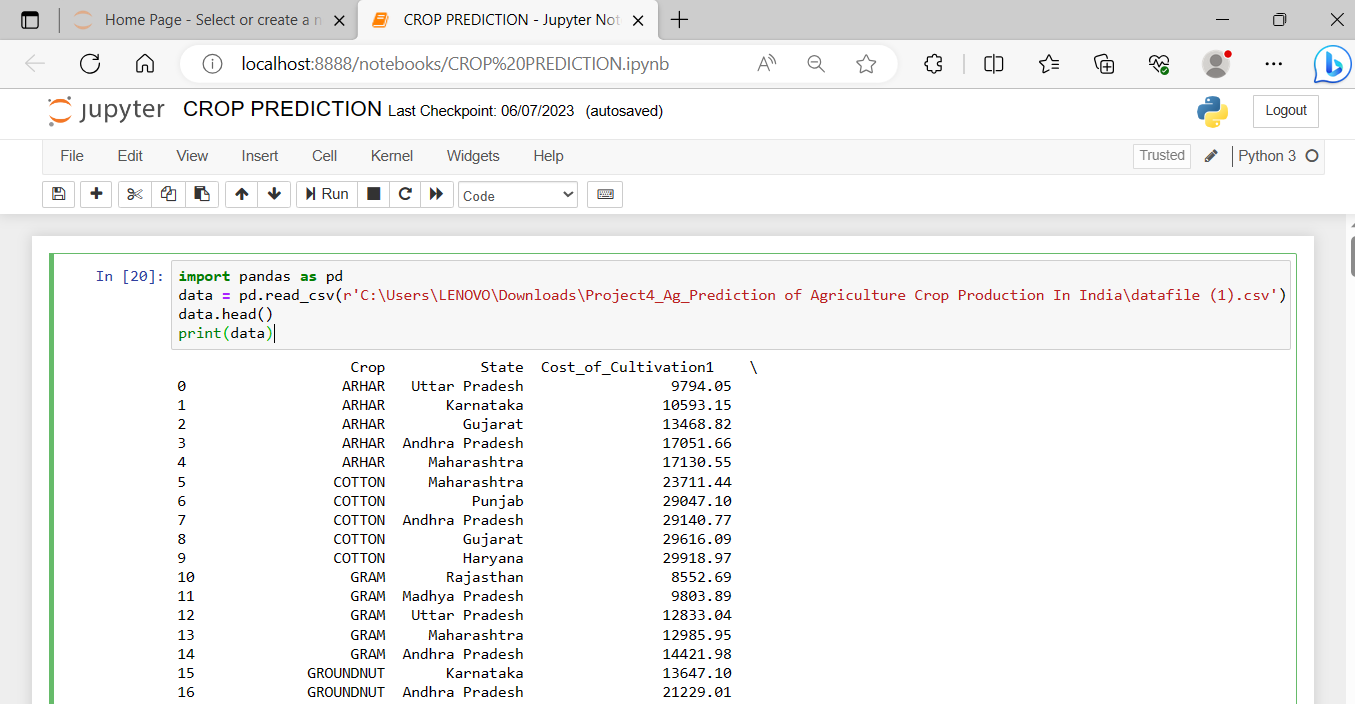
Testing

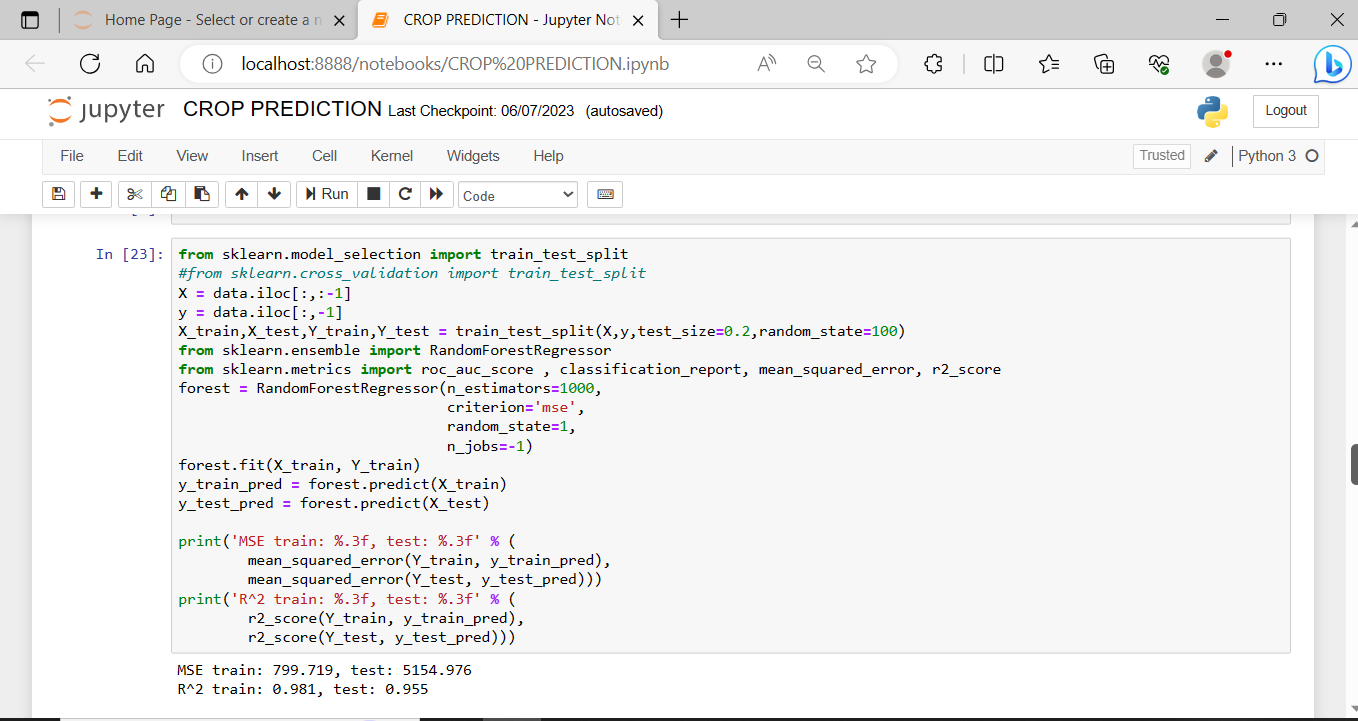
Data set

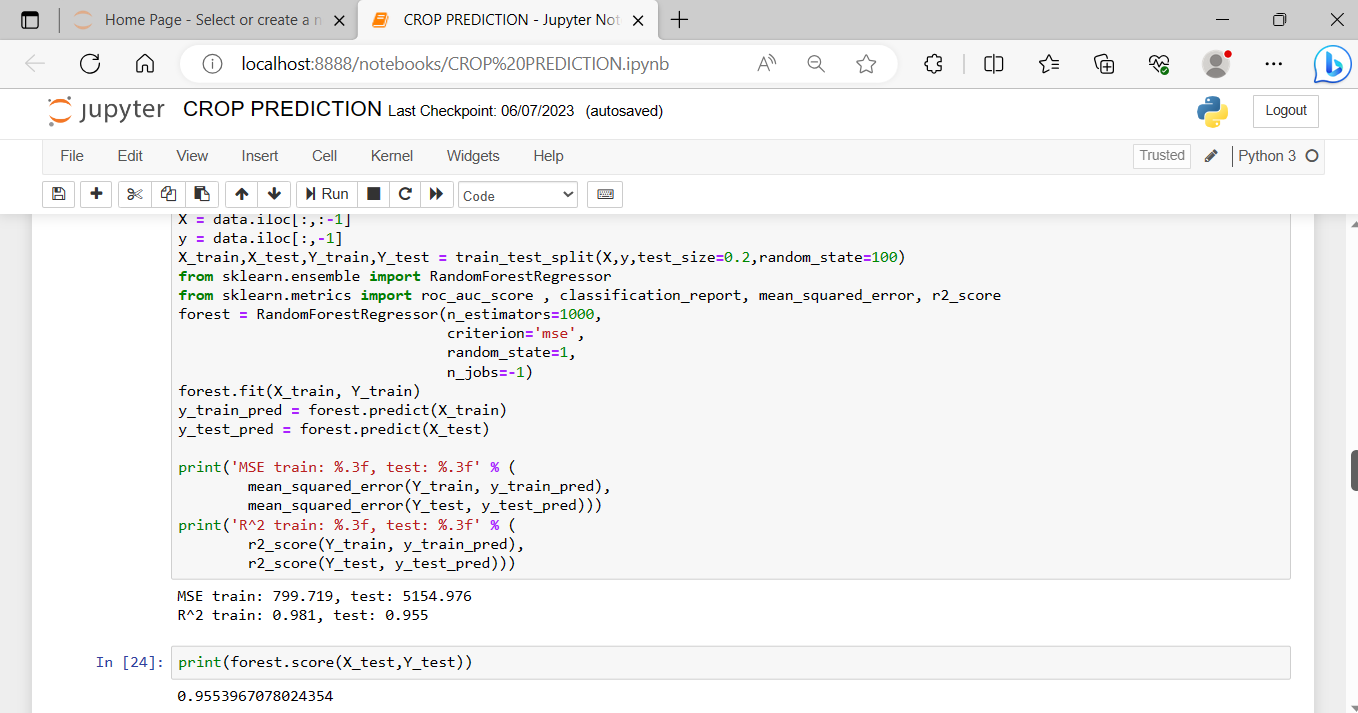
Training

Data set

* This week we implemented coding part . Here we take the datafile 1 for our random forest prediction for predicting the yield prediction .
* Here we use jupyter notebook for implementation .
* In first step we import libraries like pandas and numpy .
* Next step we read the dataset and show that data set.
* And then we need to do preprocess the data .
* So that purpose we import lable encoder from scikit learn (sklearn) and we encode the data .
* And then we import model selection train test model (basically 80 :20 ) from sklearn , and then we use random forest classifier for our model and fit the data models .
* And then find the score for our model it shows as 95.5 % is is good one .
* We attach the outputs still now we got .

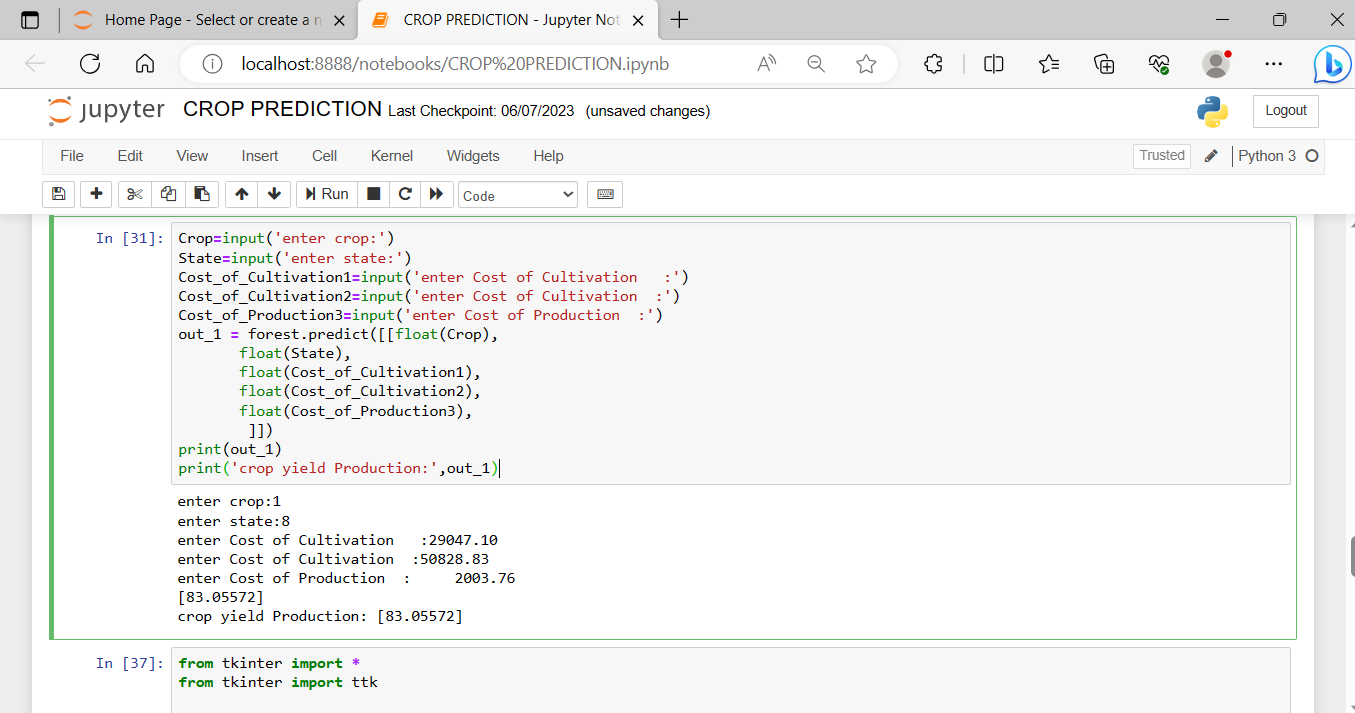


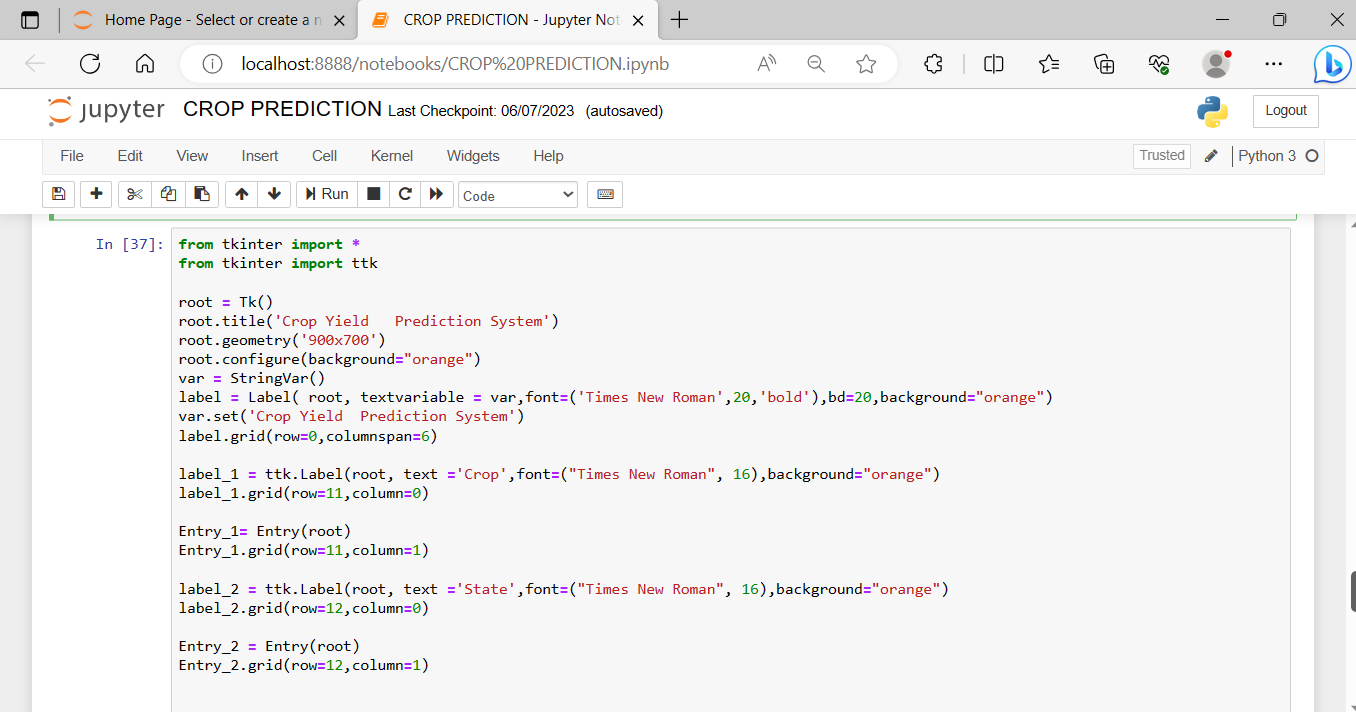


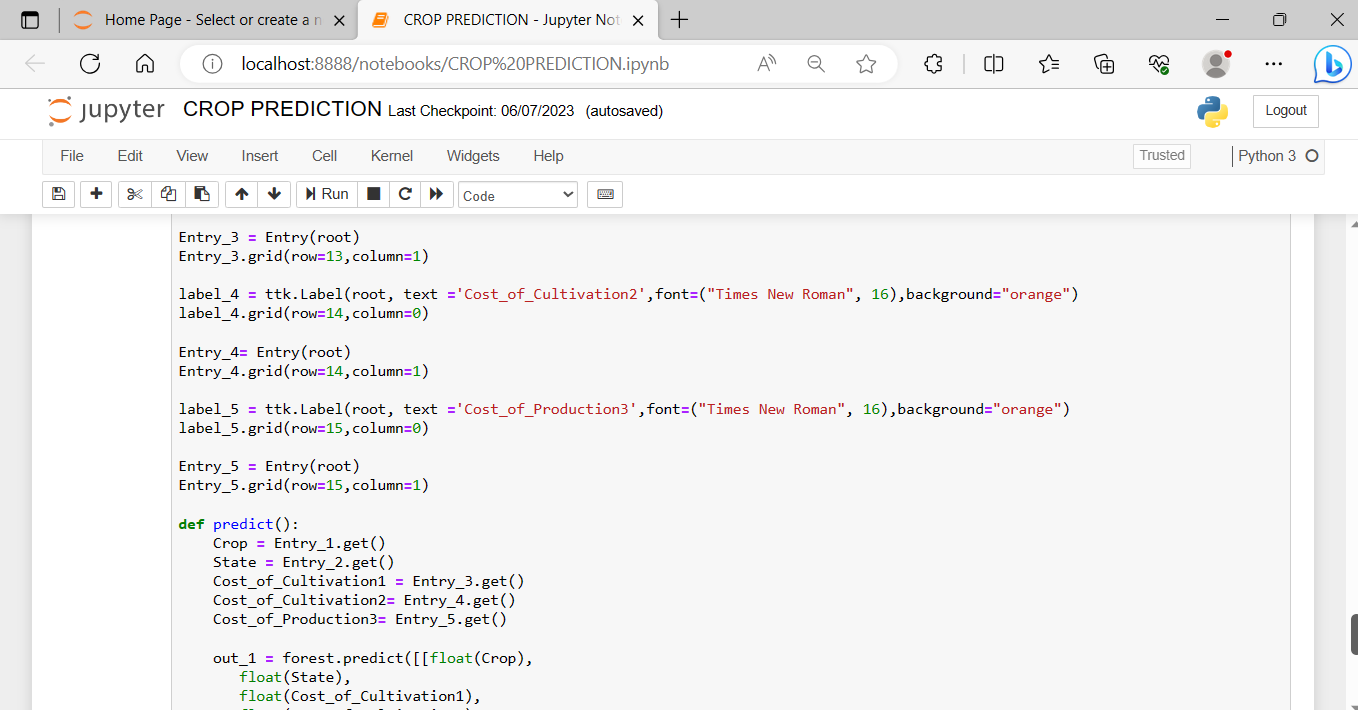


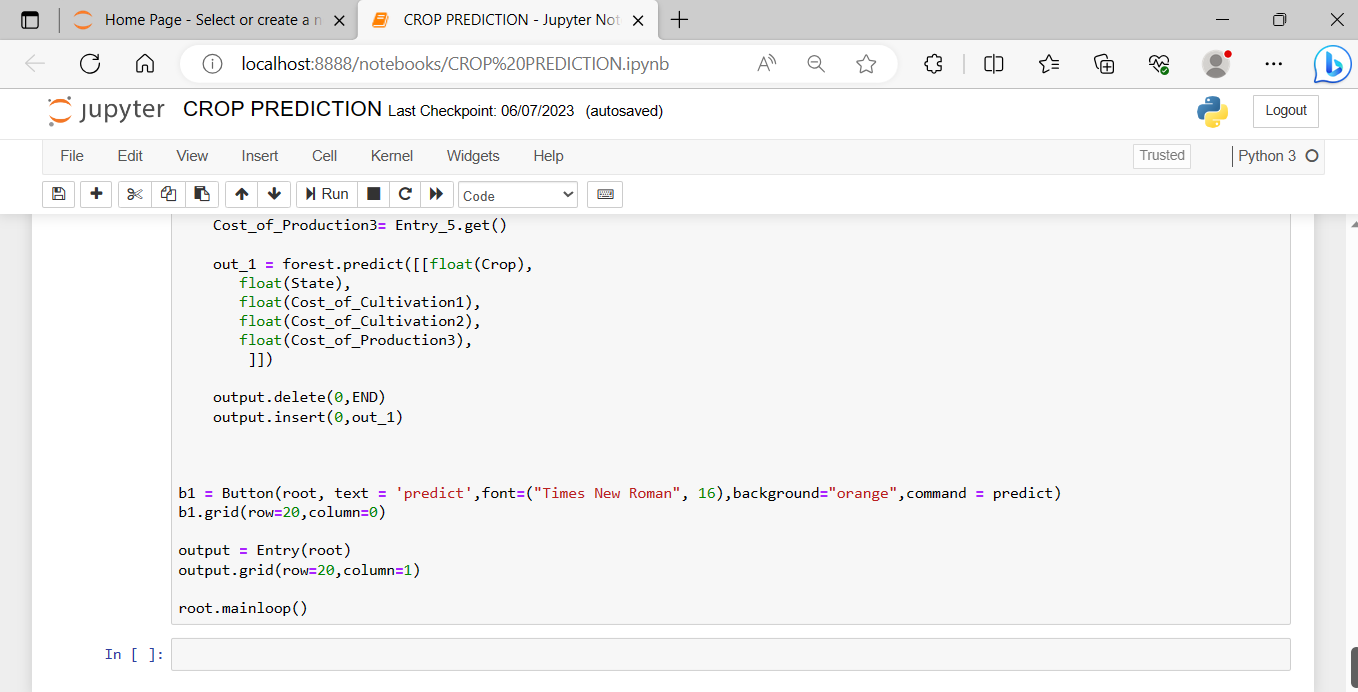
WEEK 4:

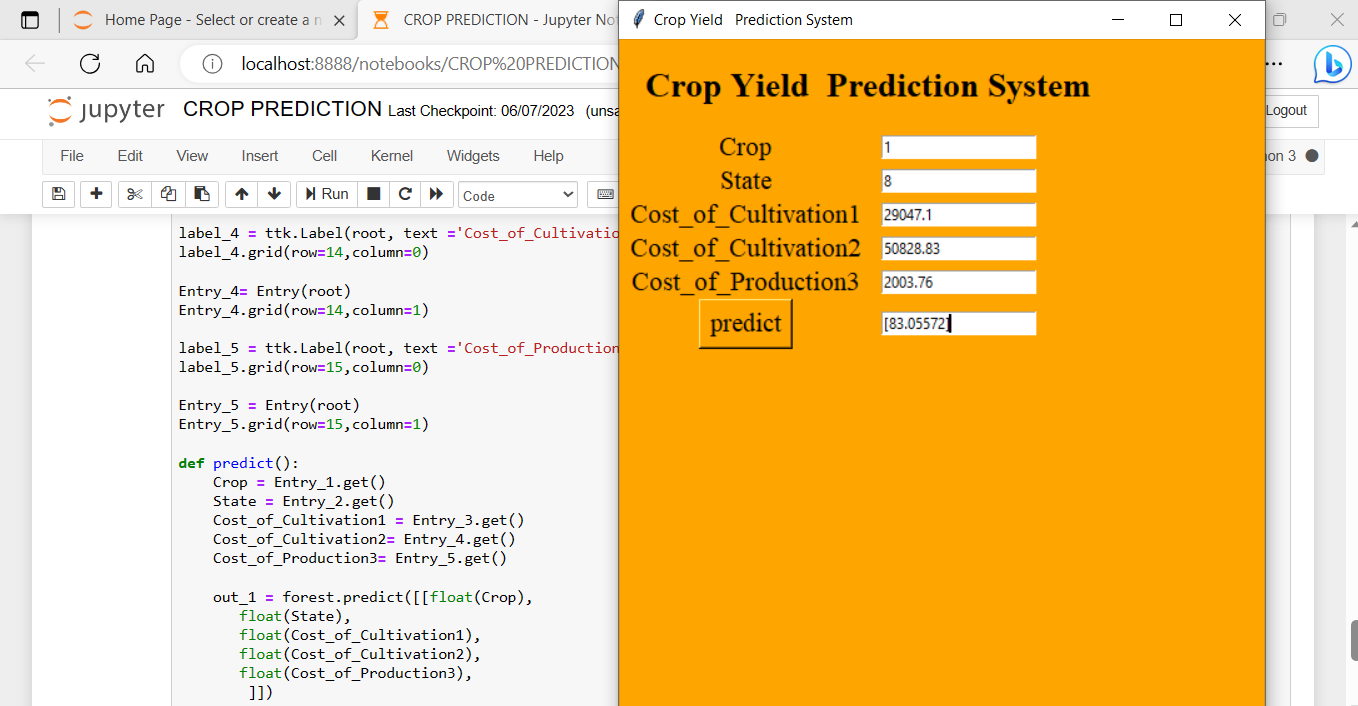
This week we completed our predicted output and our visualized page using tkinter.











WEEK 5:

This week we completed our accuracy test and mean square error and r^2 error . Here our random forest model score is 95.53 % so no need to improve our model it’s good one.

