International Journal of Modern Agriculture
ISSN: 2305-7246
Volume 10 Issue 2, 2021

Web Application Development for Site Specific Crop Prediction using Machine Learning

Sivanand K[1], Sai Amrutha M[2], Chaitra Nayak J[3]

[1]Electronics and Communication Engineering, REVA University, (INDIA)

[2]Electronics and Communication Engineering, REVA University, (INDIA)

[3]Electronics and Communication Engineering, REVA University, (INDIA)

ABSTRACT

Site-specific crop prediction (SSCP) is a crop prediction technique that is based upon observing, measuring, and responding to various aspects such as rainfall, climate, and soil conditions that affect the growth and yield of crops. It is one of many modern farming techniques which can make production more efficient. And to ensure soil conservation practices also. With the SSCP farmers can take a better yield and profit, the whole farming area can be utilized effectively, and the overall farm efficiency will also increase. In this we make use of the weather data, humidity, rainfall, land location, and other factors which affect the crop growth and yield rate to predict the best cost-effective crop using a prediction-based machine learning algorithm. The algorithm which we have used is the Random Forest Algorithm and some attributes i.e., rainfall, average maximum temperature, and average minimum temperature of a certain location, and gives prediction based on yield rate per unit area. By making use of this technique, we can produce better yield and can avoid losses.

Keywords: Food Security, Humidity, Random Forest Algorithm, Site Specific Crop Prediction.

1. INTRODUCTION

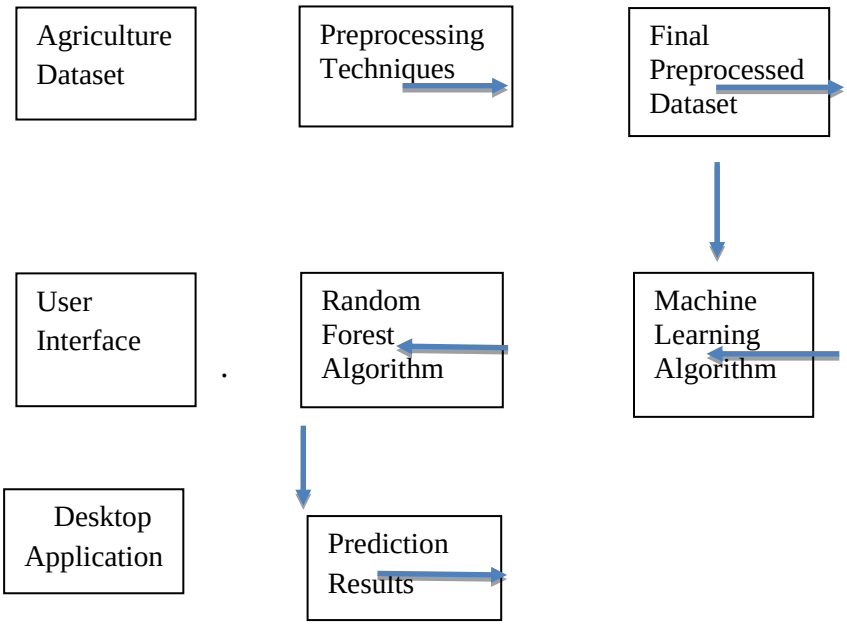
Machine Learning in agriculture aims at increasing productivity and maximizing the yield of crops. Agriculture production plays a very important factor in the Indian economy, and sufficient food supply for the current population is a big issue. The demand for supplying the crops is always constant throughout the year because food is important for humans to live. However certain hindrances could affect the supply of the crops, such as changing of climate, the growth of crops is not stable and constant. Thus, the productivity is inconsistent which leads to the fluctuating of

prices on crops. Majorly the croplands hold is of type small/marginal scale. Whereas crop production is dependent on natural parameters like soil and environment. For small/marginal scale farmers appropriate, user-friendly and low-cost system is needed. As these farmers prefer nature-dependent rain-fed cropping, those natural parameters need to be considered while taking decisions regarding the agriculture production system.

Various machine learning techniques are available which can predict the suitability level according to input and training dataset. Site-Specific Crop Prediction is a predictive application that is able to predict the most effective crop by taking input of location and planting date by farmers. Site-Specific Crop Prediction [SSCP] is popular among small agricultural farm holdings. It increases farmer profitability by increasing yield and rationalizing farm inputs. It also mitigates the detrimental effects of agricultural practices on the environment.

SSCP is performed relatively easier by the farmers in smaller farms because they have access to every inch of the land. With the growing farm size, land and crop information is sparsely accessible, hampering informed decision-making for SSCP. Consequently, farmers use uniform rate inputs on larger yields. The machine learning algorithm which has been incorporated in this application is Random Forest Algorithm. It increases the farmer's profit margin by giving the appropriate suggestion of crops. Our goal is to make a project prototype that will be easy to operate even for amateurs in technology usage. This will also help the farmer to remember different farming schedules. The entire crop cycle can benefit from an application by taking inputs from the farmers like district name, year, season, etc. Agriculture being one of the most important sectors of income in this country had far less involvement of technology.

II.METHODOLOGY



2.1Agriculture Dataset:

The first step is to collect the data. Initially the data collected from various resources like government websites based on the requirement. Here we have collected two datasets one being rainfall dataset, where rainfall dataset contains 190 tuples and 10 attributes and second dataset contains various attributes like district name, season, area and crop name and it contains 2,46,792 tuples and 7 attrib*utes.* Few rows ofboth the datasets are as shown below in Fig 1 and Fig2.

2.2DataPre-Processing:

The second step is data pre-processing. Pre- processing is mainly performed to remove unwanted and null data values from the dataset. Indirectly it is called as cleaning of data. There are many pre-processing techniques available in data mining. For our data we have used four among them and they are

1. Feature subset selection
2. Dimensionality Reduction
3. Data Aggregation
4. Data Transformation

2.2.1Feature subset selection:

It is a procedure of selecting a set of attributes which are related for a model creation. Here the Data-set initially consists data of all states in India where we have selected only Karnataka State. Because our work requires only Karnataka state as our main aim is to predict the yield and rainfall of districts in Karnataka state, so we have decreased the size of dataset from 2,46,092 tuples to 21,121 tuples. Further we have chosen only three major crops from each district in Karnataka state which further reduces the size of dataset to 2503 tuples.

2.2.2Dimensionality Reduction:

Dimensionality Reduction is the process of transformation of data from a high-dimensional space into low-dimensional space so that the low-dimensional representation retains the meaningful original data. That means it is the procedure of deleting unwanted or not required attributes in the dataset before creating the model.
Initially the dataset contains State Name attribute containing states all over India. But in our work the main goal is to predict the yield of crops of Districts within the state of Karnataka.

2.2.*3Data Aggregation:*

Data Aggregation is the process of gathering the data and presenting it in a summarized format. The data may be gathered from multiple data sources with the intent of combining these data sources into a summary for data analysis.
Here we have combined two datasets namely Crop dataset and rainfall dataset after applying dimensionality reduction. As a result, we will be adding Rainfall attribute to crop dataset.

2.2.4Data Transformation:

Data Transformation is the procedure of conversion of data from one to another format typically from the format of a source system into the required format of a destination system. In this Data Transformation is used to transform all discrete type of data into quantitative type of data, where we can apply the required regression algorithms.

International Journal of Modern Agriculture
ISSN: 2305-7246
Volume 10 Issue 2, 2021

	A	B	C	D	E	F	G
1	State_Nan	District_Na	Crop_Year	Season	Crop	Area	Production
2	Andaman a	NICOBARS	2000	Kharif	Arecanut	1254	2000
3	Andaman a	NICOBARS	2000	Kharif	Other Khar	2	1
4	Andaman a	NICOBARS	2000	Kharif	Rice	102	321
5	Andaman a	NICOBARS	2000	Whole Yea	Banana	176	641
6	Andaman a	NICOBARS	2000	Whole Yea	Cashewnu	720	165
7	Andaman a	NICOBARS	2000	Whole Yea	Coconut	18168	65100000
8	Andaman a	NICOBARS	2000	Whole Yea	Dry ginger	36	100
9	Andaman a	NICOBARS	2000	Whole Yea	Sugarcane	1	2
10	Andaman a	NICOBARS	2000	Whole Yea	Sweet pota	5	15
11	Andaman a	NICOBARS	2000	Whole Yea	Tapioca	40	169
12	Andaman a	NICOBARS	2001	Kharif	Arecanut	1254	2061
13	Andaman a	NICOBARS	2001	Kharif	Other Khar	2	1
14	Andaman a	NICOBARS	2001	Kharif	Rice	83	300
15	Andaman a	NICOBARS	2001	Whole Yea	Cashewnu	719	192
16	Andaman a	NICOBARS	2001	Whole Yea	Coconut	18190	64430000
17	Andaman a	NICOBARS	2001	Whole Yea	Dry ginger	46	100
18	Andaman a	NICOBARS	2001	Whole Yea	Sugarcane	1	1
19	Andaman a	NICOBARS	2001	Whole Yea	Sweet pota	11	33
20	Andaman a	NICOBARS	2002	Kharif	Rice	189.2	510.84
21	Andaman a	NICOBARS	2002	Whole Yea	Arecanut	1258	2083
22	Andaman a	NICOBARS	2002	Whole Yea	Banana	213	1278
23	Andaman a	NICOBARS	2002	Whole Yea	Black pepp	63	13.5
24	Andaman a	NICOBARS	2002	Whole Yea	Cashewnu	719	208
25	Andaman a	NICOBARS	2002	Whole Yea	Coconut	18240	67490000
26	Andaman a	NICOBARS	2002	Whole Yea	Dry chillies	413	28.8
27	Andaman a	NICOBARS	2002	Whole Yea	Dry ginger	47.3	133
28	Andaman a	NICOBARS	2002	Whole Yea	Sugarcane	5	40
29	Andaman a	NICOBARS	2003	Kharif	Rice	52	90.17

crop_production (2)

Fig 1.Crop Dataset

District	Year	Kharif_rainfall	Rabi_rainfall	summer_rainfall	WholeYear_rainfall
Bagalkote	2013	77.075	0.6	59.85	45.84166667
Bagalkote	2014	97.475	0.275	41.25	46.33333333
Bagalkote	2015	52.975	3.475	50.65	35.7
Bagalkote	2016	114.525	1.075	62.875	59.49166667
Bagalkote	2017	71.05	11.05	47.125	43.075
Bagalkote	2018	141.175	15	50.75	68.975
Bagalkote	2019	90.35	24.275	44.425	53.01666667
Bangalore Rural'	2013	170.25	8.875	92.3	90.475
Bangalore Rural'	2014	194.525	27.525	110.825	110.9583333
Bangalore Rural'	2015	47.375	20.35	60.7	42.80833333
Bangalore Rural'	2016	146.15	10.05	45	67.06666667
Bangalore Rural'	2017	156.6	7.4	57.15	73.71666667
Bangalore Rural'	2018	103.2	12.825	54.425	56.81666667
Bangalore Rural'	2019	122.25	40.075	79.075	80.46666667
Bangalore Urban'	2013	164.95	8.075	83.5	85.50833333
Bangalore Urban'	2014	237.475	17.2	83.025	112.5666667
Bangalore Urban'	2015	44.55	12.15	103.075	53.25833333
Bangalore Urban'	2016	171.8	13.95	60.4	82.05
Bangalore Urban'	2017	186.6	14.475	73.025	91.36666667
Bangalore Urban'	2018	140.725	21.1	82.075	81.3
Bangalore Urban'	2019	119.775	43.725	76.525	80.00833333
Belgam	2013	120.5	1.775	80.65	67.64166667
Belgam	2014	217.45	1.525	79.725	99.56666667
Belgam	2015	165.675	10.25	78.075	84.66666667
Belgam	2016	150.4	5.475	80.35	78.74166667
Belgam	2017	125.275	8.475	66.575	66.775
Belgam	2018	249.2	16.275	54.85	106.775
Belgam	2019	168.075	27.05	56.725	83.95

Fig 2. Rainfall Dataset

B	C	D	E	F	G
District Na	Crop_Year	Season	Crop	Area	Productior
BAGALKOT	2009	Kharif '	Maize	37419	130280
BAGALKOT	2009	Kharif '	Sunflower	51307	29116
BAGALKOT	2009	Rabi '	Maize	12094	42223
BAGALKOT	2009	Rabi '	Sunflower	62717	35641
BAGALKOT	2009	Summer	Maize	1509	5691
BAGALKOT	2009	Summer	Sunflower	4087	3052
BAGALKOT	2009	Whole Yea	Onion	6986	65436
BAGALKOT	2010	Kharif '	Maize	38948	153176
BAGALKOT	2010	Kharif '	Onion	6732	39517
BAGALKOT	2010	Kharif '	Sunflower	48722	26039
BAGALKOT	2010	Rabi '	Maize	14743	61920
BAGALKOT	2010	Rabi '	Onion	1478	15322
BAGALKOT	2010	Rabi '	Sunflower	81668	56526
BAGALKOT	2010	Summer	Maize	1723	7038
BAGALKOT	2010	Summer	Onion	675	9735
BAGALKOT	2010	Summer	Sunflower	3328	3098
BAGALKOT	2011	Kharif '	Maize	36326	137679
BAGALKOT	2011	Kharif '	Sunflower	40031	11199
BAGALKOT	2011	Rabi '	Maize	12637	44443
BAGALKOT	2011	Rabi '	Sunflower	67538	20690
BAGALKOT	2011	Summer	Maize	2128	6625
BAGALKOT	2011	Summer	Sunflower	3919	3902
BAGALKOT	2011	Whole Yea	Onion	8284	59102
BAGALKOT	2012	Kharif '	Maize	49303	194845
BAGALKOT	2012	Kharif '	Onion	12075	81411
BAGALKOT	2012	Kharif '	Sunflower	22559	13004
BAGALKOT	2012	Rabi '	Maize	16147	67418
BAGALKOT	2012	Rabi '	Onion	971	10215

Fig 3: Final Dataset

III.ALGORITHMS

Machine Learning Algorithms are categorized as

1) Classification

2) Regression

3) Association

4) Clustering

Supervised learning and Unsupervised learning are the types of Machine Learning techniques. Supervised learning provides a labelled dataset whereas Unsupervised learning provides unlabeled data. As we require a labeled dataset so we have selected Supervised Learning. Regression and Classification are the types of Supervised Learning and Association and clustering are the types of Unsupervised Learning.

Here we are using Supervised learning technique and also our data is a continuous one so we have used regression. Regression is a data mining technique that is used to model and analyze the

relationship between variables and often times how they contribute and are related to producing a particular outcome together.

In regression we have chosen Random Forest Algorithm. Random Forest Algorithm is a set of decision trees where each decision tree predicts the output and finally selects the output which has more accuracy.

```
import pandas as pd
import numpy as np

from sklearn import linear_model

crop = pd.read_csv("D:/MajorProject1/crop_num1.csv")
crop.head()
```

	District_Code	Crop_Year	Season_Code	Rainfall	Area	Crop_Code	Production
0	4.0	2009	0.0	92.08	37419	4.0	130280.0
1	4.0	2009	0.0	92.08	51307	8.0	29116.0
2	4.0	2009	1.0	7.96	12094	4.0	42223.0
3	4.0	2009	1.0	7.96	62717	8.0	35641.0
4	4.0	2009	2.0	50.98	1509	4.0	5691.0

```
X=crop[['District_Code','Crop_Year','Season_Code']]
y=crop['Rainfall']

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=0)

from sklearn.ensemble import RandomForestRegressor

regressor = RandomForestRegressor(n_estimators=20, random_state=0)
regressor.fit(X_train, y_train)
y_pred = regressor.predict(X_test)
```

Fig 4: Python code to predict the output

Our Python programming code is written in Jupyter Notebook and executed. Jupyter is a free, open-source, interactive web tool known as a computational notebook, which researchers can use to combine software code, computational output, explanatory text and multimedia resources in a single document.

In the above code some of the libraries like pandas, NumPy, sklearn are used. Pandas is a software library written for Python programming language. In pandas “import pandas as pd” can present data in a way that is suitable for data analysis via its series and Data Frame structures.

NumPy is a library for the python programming language. It provides a high-performance multidimensional array object, and tools for working with these arrays. Sklearn(scikit-learn) is a free software machine learning library for the Python Programming Language.

A. Desktop App

We have used Python 3.6.9 as a platform to develop the user-interface code and converting it into .exe file where we can install the .exe file as a Desktop App. The below code is a part of User Interface Code.

```
from tkinter import *
import tkinter as tk
import PIL.Image
from PIL import ImageTk
from tkinter import ttk
import pickle
from tkinter.ttk import Progressbar
import time

windo = Tk()
windo.configure(background='white')
windo.title("Site Specific Crop Prediction")
windo.geometry('820x570')
windo.iconbitmap('./Meta/crop.ico')
windo.resizable(0,0)
rain_model = pickle.load(open('./Model/Rainfall_Model.pkl', 'rb'))
production_model = pickle.load(open('./Model/Production_Model.pkl', 'rb'))


dist_list = ['Bangalore_Rural','Bangalore_Urban','Dakshin_Kannad','Uttar_Kannad','Bagalkot',
             'Belgaum','Bellary','Bidar','Bijapur','Chamrajanagar','Chikmagalur','Chitradurga',
             'Davangere','Dharwad','Gadag','Gulbarga','Hassan','Haveri','Kodagu','Kolar',
             'Koppal','Mandya','Mysore','Raichur','Shimoga','Tumkur','Udupi']

dist_list_cp = dist_list.copy()
dist_list_cp.insert(0,'--Select--')

year_list = [2008,2009,2010,2011,2012,2013,2014,2015,2016,2017,2018,2019,2020,
             2021,2022,2023,2024,2025,2026,2027,2028,2029,2030]
year_list_cp = year_list.copy()
year_list_cp.insert(0,'--Select--')
```

Fig 5: Python code for user interface

The above code is part of the Python code that was built for user interface which was shown in Fig:6.”Tkinter” is a standard GUI (Graphical User Interface)library for Python. Python when combined with Tkinter provides a fast and easy way to create GUI applications.Tkinter provides a powerful object-oriented interface to the Tk GUI tool kit.PIL.Image is the Python library which provides the python interpreter with image editing capabilities. Python Pickle module is used for serializing and de-serializing python object structures. The process to convert any kind of python objects(list,dict)into byte streams is called pickling or serialization.

B.Conversion from GUI to .EXE file:

To convert from GUI to .EXE file we need a converter using a graphical interface and Pyinstaller in Python .As a Prerequisite we need Python version above Python 3.5 as it cannot support below version 3.5.We can install the project using “$ pip install auto-py-to-exe”. After installing the project we have to run the following in the terminal.

"$ auto-py-to-exe."

In the App the user will be asked to enter the District Name, Year of cultivation, Season, Area of land in hectares in which he is going to cultivate and the Crop Name. The App will be predicting the Rainfall in that Year and the Production in Tones.

IV.RESULTS AND DISCUSIONS

A Python language code is written using machine learning technique called Random Forest Algorithm in Jupyter Notebook. The code is processed in Jupyter Notebook. An output is obtained as shown in figure below.

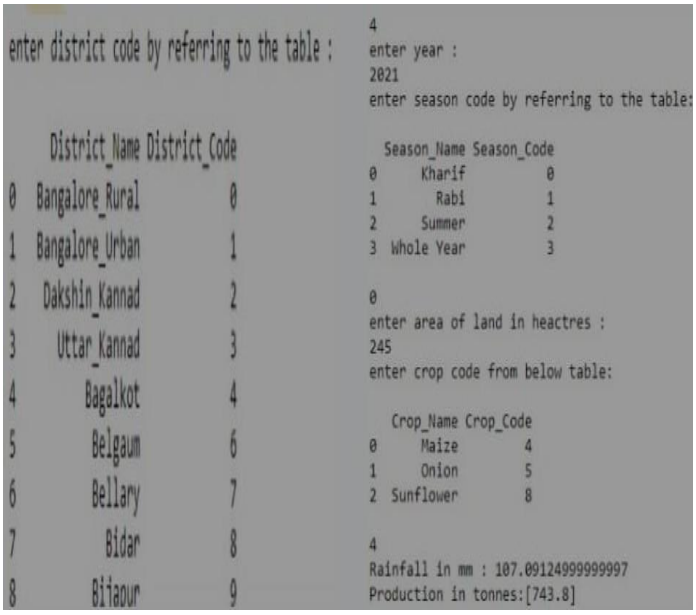


Fig 6: output of the jupyter notebook

In the above figure it will ask the user to enter the district code and after entering the district code the user has to enter the year of cultivation and season code he wants to cultivate. Alsohe has to enter the area of Hectares he wants to cultivate. After entering all this data the user needs to select the crop that he wants to cultivate in his land. By taking all this attributes as inputs the code will process the data and gives the output as the Rainfall in that year and Production in tones for the area of land he wants to cultivate.

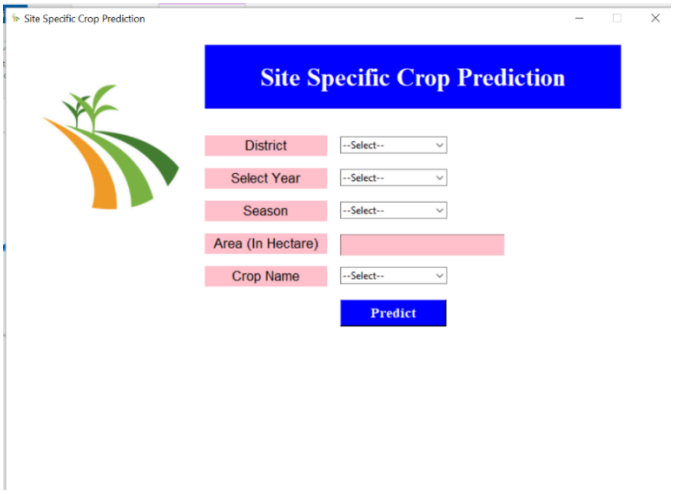


Fig 7: User Interface of the Desktop App

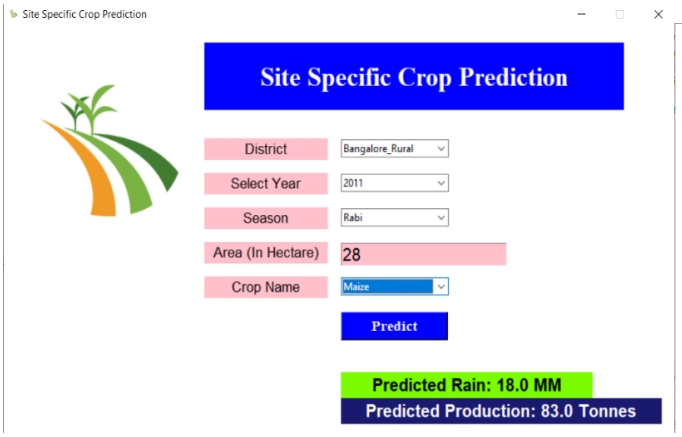


Fig 8: Bangalore Rural District Crop Prediction

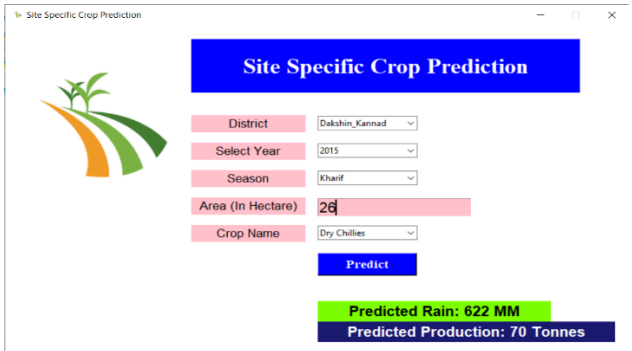


Fig 9:Dakshin Kannada District Crop Prediction

The above figure is the complete user interface of the Desktop App where the user has entered the District, Year, Season, Area in Hectares and also the Crop Name. After entering all this attributes, the App will predict the output as Rainfall in the Year he wants to cultivate and production in tonnes as shown in above figure where we have entered District as Banglore_Rural, year as 2011, Season as Rabi, Area of hectares as 28 and crop name as Maize. The output we have got is 18.0MMRainfall in the year 2011 and Predicted Production in tonnes is 80.0Tonnes for 28 acres of land.

V.CONCLUSION

As we discussed earlier, predicting the crop yield and rainfall in that year before cultivating it has lot advantages like farmer can plan about his investment and can cultivate accordingly without any losses to some extent. So, we had built a machine learning model and developed it into a Desktop App to predict the yield of crops. Initially datasets are collected from different resources and appropriate preprocessing techniques were applied to obtain final dataset. Different regression algorithms were applied and it was clearly found that Random Forest algorithm was more précised and also accurate when compared to other regressor algorithms. A model was built in jupyter notebook using python and the output was obtained. After that python code is written in PyCharm GUI file is obtained. Then GUI file is converted into .exe file for a better user-interface where farmer can enter the required inputs and can obtain the predicted rainfall in the year that he mentioned and yield of crop.

VI.FUTURE ENHANCEMENT

In future an android Application with more features, user interface will be developed and uploaded in play store.

VII.REFERENCES

[1].B.Manjula Josephine, K.Ruth Ramya, K.V.S.N Rama Rao, Swarna Kuchibhotla, P. Venkata Bala Kishore and S. Rahamathulla "Crop Yield Prediction Using Machine Learning", International Journal Of Scientific & Technology Research Volume 9, Issue 02, February 2020.

[2]. Sangeeta and Shruthi G "Design And Implementation Of Crop Yield Prediction Model In Agriculture", International Journal Of Scientific & Technology Research Volume 8, Issue 01, January 2020.

[3]. RatnmalaNivruttiBhimanpallewar And Manda Rama Narasingarao, "Alternative Approaches Of Machine Learning For Agriculture Advisory System", 10th International Conference On Cloud Computing, Data Science & Engineering (Confluence) IEEE 2020.

[4]. Aruvansh Nigam, Saksham Garg, Archit Agrawal and Parul Agrawal "Crop Yield Prediction Using Machine Learning Algorithms", 2019 Fifth International Conference On Image Information Processing (Iciip).

[5] ACCLIMATIZATION OF RICE (ORYZA SATIVA) UNDER WATERSTRESS, Deeksha Prakash Singh, International Journal Of Advance Research In Science And Engineering http://www.ijarse.com IJARSE, Volume No. 10, Issue No. 01, January 2021 ISSN-2319-8354(E).
[6]. Vanitha Cn, Archana N and Sowmiya R "Agriculture Analysis Using Data Mining And Machine Learning Techniques", 2019 5th International Conference On Advanced Computing & Communication Systems (ICACCS).
[7]. Meeradevi and Hrishikesh salpekar "Design And Implementation Of Mobile Application For Crop Yield Prediction Using Machine Learning", 2019 Global Conference For Advancement In Technology (GCAT) Bangalore, India. Oct 18-20, 2019.
[8].R.Karthikeyan ,M.Gowthami,A.Abhishek, P.Karthikeyan Implementation of Effective Crop Selection By using The Random Forest Algorithm International Journal of Engineering & Technology .
[9]. Md Tahmad Shakoor, Krishna Rahman, Sumaiya Nasrin Rayta,Amitabha Chakrabarty Agricultural Production output predictions using Supervised Machine Learning Technique