International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 08 | Aug 2020 www.irjet.net p-ISSN: 2395-0072

Forecasting Platform for the Farmers

Premjit Kumar

Student, School of Electronics Engineering, KIIT DU, Bhubaneswar, Odisha, India

---

Abstract - *Agriculture is principal means of livelihood for most of India. The sector provides livelihood to about 50% of the country's workforce. Weather phenomenal like rainfall, temperature and many more plays a vital role in the production of agricultural sector. As it is said too much or too little can have its adverse effect on the field of production. But modern technology tools such as proper weather forecasting could not reach the rural areas which results in the decrease of productivity and the product quality as well. The productivity of agriculture can be measured in terms of the ratio of agricultural inputs with respect to the agricultural outputs. The individual products are normally measured by weights, but their varying densities makes the measurement of overall agricultural outputs very difficult. Therefore, the output can be measured as the market value of the final outputs, which do not include the intermediate products like corn feeds. The Crop Production Index or the CPI shows the production of the agriculture of each year with respect to its the base period. This process includes all crops except the fodder crops. Regional and income group aggregates for the FAO's production indexes are calculated from the underlying values in international dollars, normalized to the base period. This project aids to analyze the different types of crop grown all over the country, its proper analysis & prediction of the amount of rainfall in different regions of the country and finally predicting the Crop Production Index Values of different years.*

Key Words: **Agriculture, Crops, Prediction, Machine Learning, Random forest regression, SVM, ANN, Linear Regression.**

1. INTRODUCTION

Agriculture is the science and art of cultivating plants and livestock. In the rise of human civilization, agriculture plays a very vital role. Farming of domesticated species fulfils the demand of food which enabled the people to live in the cities and improvise the technologies. Agriculture is the oldest sector of the country and it also can be called as the backbone of the nation as the entire nation is totally dependent in this sector. So, to develop this industry, it has been a great concern for the country, population growth, food security concern, climate change and many more like factors have driven this sector to seek more creative approaches to improvise the crop yielding techniques and hence to produce a better farming result. In agriculture sector, the farmers and agro based industries have to take several decisions every day and there are various factors that influence them. Some of the factors on which agriculture depends are soil, climate, cultivation, irrigation, fertilizers, temperature, rainfall, harvesting and use of pesticides. Now a days, farmers can get regular updates regarding agricultural information like proper crop selection, required pesticides and fertilizers, weather forecating and many more information from various resources like radio, television, newspapers etc. which are distributed on many locations. But the problem is the data containing various important updates may be heterogeneous in structure. Therefore, it is required to develop a platform where the farmers can get access to the required information directly. Here comes the role of AI.

Artificial Intelligence being a game changer in several other industries can start to leverage this technology in developing the sector.

2. BRIEF DESCRIPTION

The science of training machines to learn and produce models for future predictions is widely used, and not for nothing. Agriculture plays a critical role in the global economy.
In the project I used several machine learning techniques to extract information from agriculture as well as rainfall data and to give proper visualization regarding the crops, the amount of rainfall and make future predictions so that agriculture can be carried out in a planned manner.

The main objectives of the project includes-

- 1. Analyzing the different types of crop grown all over the country
- 2. Proper analysis & Prediction of the amount of rainfall in different regions of the country
- 3. And finally predicting the Crop Production Index Values of different years.

The software used to perform all the analysis during the project is Python 3 based on Jupyter notebook of Conda distribution.

3. DATASET USED

In the field of Machine Learning, the datasets play a vital role to perform any kind of analysis or prediction in a program.

In the project, I used the following datasets:

a. Apy Dataset: Better known as crop production statistics. This dataset contains all the state names, year & season of production including the crops grown; area used by each

crop & finally the amount of production of the crop in every year. [3]

b. Rainfall Dataset: Better known as Rainfall Statistics. It contains the name of all the states, amount of rainfall in different months, the annual & the seasonal rainfall data starting from the year 1901 - 2015. [1]

c. District Wise Dataset: This dataset contains all the state names as well as all the district names of the that state & rest of the data are similar to the previous dataset. [2]

All the datasets were downloaded from the site data.gov.in, containing all genuine and relevant data.

The project shows analysis, visualization and prediction of different data according to the requirement.

4. PROBLEM STATEMENT

4.1. Rainfall Analysis

A. Background Theory

All plants need a proper amount of water to survive, therefore rain which is the the most effective way of watering is important to the agricultural field. While a regular rain pattern is usually vital to healthy plants but too much or too little rainfall can be devastating to crops. Drought can kill crops and increase erosion, while overly wet weather can cause harmful fungus growth. Rain is usually seen as a benefit to crops and fields, but there is an "ideal" amount of rainfall in any given growing season for most crops. If the average rainfall is much lower or higher than the ideal, it can lead to significant problems, from drowned crops to lower yields. Rainfall can also determine how fast a crop will grow from seed, including when it will be ready for harvesting. A good balance of rain and proper irrigation can lead to faster-growing plants, which can cut down on germination time and the length between seeding and harvest. The crops are dependent on water during their entire lifecycle in order to survive and thrive. [16]

Soil is also greatly affected by rainfall. Additionally, as mentioned previously, overwatering or too much rain can also lead to bacteria, fungus, and mold growth in the soil.

B. Problem Statement

Agriculture in India is largely depending on monsoon. As a result, production of food-grains fluctuates year after year. Heavy rainfall prediction is a major problem for meteorological department as it is closely associated with the economy and life of human. Accuracy of rainfall forecasting has great importance for our country because Indian economy is largely dependent on agriculture. Due to dynamic nature of atmosphere, statistical techniques fail to provide good accuracy for rainfall forecasting.[8]

4.2. The Crops

A. Background Theory

A crop is a plant that is cultivated or grown on a large scale. In general, crops are grown so they can be commercially traded. In other words, a crop is any plant that is grown and harvested extensively for-profit purposes.

India is the top producer country of many crops.

India is the largest producer of bananas and mangoes in the world? Also, it is the second-largest producer of wheat and rice. In fact, Agriculture has always been the backbone of our country's economy. Ever since the Green Revolution, I have started cultivating a variety of crops.

The major crops in India can be divided into 4 categories:

- **Food Grains**, that contains Rice, Wheat, Maize, Millets & Pulses.
- **Cash Crops**, which contains Cotton, Jute, Sugarcane, Oilseeds etc.
- **Plantation Crops**, like Tea, Coffee, Coconut & Rubber.
- **Horticulture Crops**, such as Fruits & Vegetables.

On the basis of season, the crops can be divided into:

Ravi, Kharif, Summer, Autumn, Winter & Whole Year. [6]

B. Problem Statement

In recent years there has occurred a fall in agricultural production mainly due to fall in the output of non-food articles. Moreover, rabi production has become as important as kharif production in the late 1990s. In 1999-2000, for example, of the total grain production of 209 mn. tones, rabi accounted for 104 mn. tones. This indicates a structural change in agricultural production. [9]

4.3. Crop Production Index

A. Background Theory

Agricultural productivity is measured as the ratio of agricultural outputs to agricultural inputs. While individual products are usually measured by weight, their varying densities make measuring overall agricultural output difficult. Therefore, output is usually measured as the market value of final output, which excludes intermediate products such as corn feed used in the meat industry. This output value may be compared to many different types of inputs such as labor and land (crop yield). These are called partial measures of productivity.

The Crop Production Index shows the agricultural production for each year relative to the base period. It

includes all crops except fodder crops. Regional and income group aggregates for the FAO's production indexes are calculated from the underlying values in international dollars, normalized to the base period. [7]

CPI is calculated by the formula,

Crop Production Index = Production / Area.

B. Problem Statement

One of the major causes of low income of the Indian farmers is the difficulty in marketing their crops. Due to the small size and scattered nature of agricultural holdings, the productivity per acre is low. Consequently, the collection of these surpluses for the purpose of marketing presents a serious problem.

5. PROPOSED APPROACH

5.1. Rainfall Analysis

Monsoon prediction is clearly of great importance for India. Here I predicted the rainfall pattern for a long-term basis. My main motive is to predict the amount of rainfall for a particular state and district well in advance. Here I used the past data to train and predict the amount of rainfall.

5.1.1. State wise Data Analysis

Implementation:

A. Importing the Module

To perform all the operations, I first import the following modules.

- NumPy
- Pandas
- Matplotlib
- Seaborn
- Sklearn

B. Loading the Dataset

Here I used Rainfall dataset to collect all the information.

C. Analysing the Dataset

As mentioned earlier, the dataset contains the name of all the states, amount of rainfall in different months & the annual the seasonal rainfall data starting from the year 1901 to 2015.

D. Data analysis

Then different graphs are plotted to analyze the amount of rainfall in different months for different regions of the country. It is observed that there is a increase in amount of rainfall over months July, August, September. To ensure our observation, then I plot the heatmap for the given data. Heat Map shows the co- relation(dependency) between the amounts of rainfall over months. From above it is clear that if amount of rainfall is high in the months of July, August, September then the amount of rainfall will be high annually. It is also observed that if amount of rainfall in good in the months of October, November, December then the rainfall is going to be good in the overall year.

E. Data Prediction

For prediction the data were formatted in a way, given the rainfall in the last three months then I try to predict the rainfall in the next consecutive month.

For all the experiments I used 80:20 training and test ratio.

- Linear regression
- SVR (Support Vector Machine - Regression)
- Artificial neural nets

Testing metrics: I used Mean absolute error to train the models. I also shown the amount of rainfall actually and predicted with the histogram plots. The training was done on two types, once training on complete dataset and other with training with only Odisha data. All means standard deviation observations are written, first one represents actual value, second one represents predictions.

5.1.2. District wise Data Analysis

Implementation:

Similarly, as above the number of attributes is same. The amount of rainfall in mm for each district is added from 1950-2000. I analyze the data individually for the state Odisha.

5.2. Crop Data Visualization

My approach is to estimate the yield of crops in different years in every state and district all over the country.

Implementation:

A. Importing the Module

To perform all the operations, I first import the following modules

- NumPy
- Pandas
- Matplotlib
- Seaborn

Description of the modules are already discussed earlier.

B. Loading the Dataset

Here I use the Apy dataset to collect all the information.

C. Analysing the Dataset

As discussed earlier, this dataset contains all the State Names, Year & Season of Production including the Crops grown, Area Used by each crop & finally the amount of Production of the crop in every year.

D. Data Processing

After analysing the data set, I then group the data available for a specific search. This could help the user to visualize his/her exact needs. The user can see the type of crops grown, the amount of production and the total areas involved in the production for a particular crop in each year.

I have grouped the data in the following ways:

- Overall grouping of the data
- State wise grouping the data
- District wise grouping of the data
- Season wise grouping of the data

This could help the user to compare the data of each state as well as district also. The user can also check the production of crop in a specific state or district for a particular year.

5.3. Crop Production Index

I aid to predict and analyze the Crop production index (CPI) value for upcoming years using machine learning techniques. I used Random Forest Regressor algorithm for that which is an optimization technique and decision trees was modelled.

Implementation:

A. Importing the Module

I used the same modules as discussed previously.
B. Loading the Dataset

I use the same Apy dataset in this program.

C. Analysing the Dataset

To find the CPI value for a year, I require only 3 things

- Year
- Production
- Areas used for the production

So out of the entire dataset I consider the data used only in this three columns and rows to make our analysis and prediction.

D. Data Processing

After grouping the data, I calculated the value of CPI by the formula discussed earlier. Then I use the boxplot to remove all the outline values. Then I analyse the variation in area and production.

Then I use Random Forest Regression Algorithm to perform our prediction. I first split the values into test and train, and then I predict the test values.

6. RESULT ANALYSIS

6.1. Rainfall Analysis

A. State wise Prediction Observation:

a. Training on complete dataset

Algorithm MAE

- Linear Regression = 96.32435
- SVR = 127.160061
- Artificial neural nets = 90.0313595

b. Training on Odisha dataset only

Algorithm MAE

- Linear Regression = 98.24317599
- SVR = 135.06085124
- Artificial neural nets = 78.355181485

It is observed that Neural Networks performs better than SVR and LR. Observed MAE is very high which indicates machine learning models won't work well for prediction of rainfall.

Odisha data has a single pattern that can be learned by models, rather than learning different patterns of all states. So has high accuracy.

c. Plots

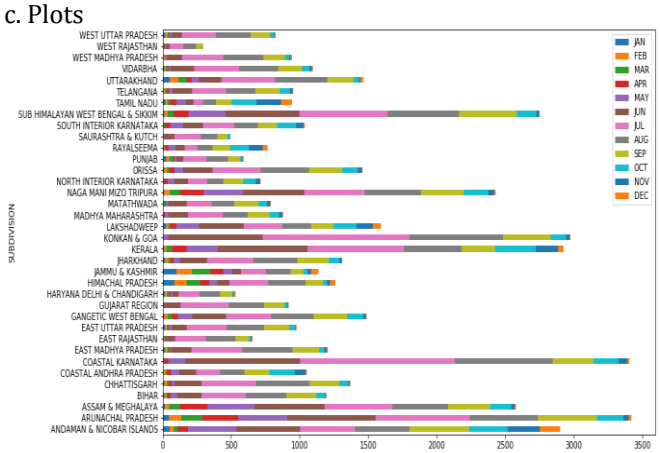

Fig i. Month wise rain in all states

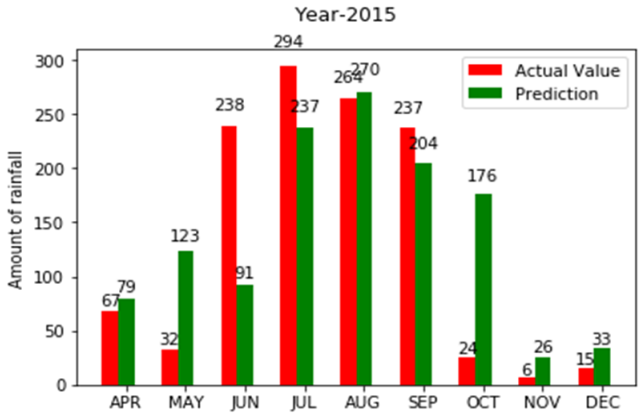

Fig ii. Plot of Actual vs Predicted values of Rainfall In Odisha 2015 for the complete data

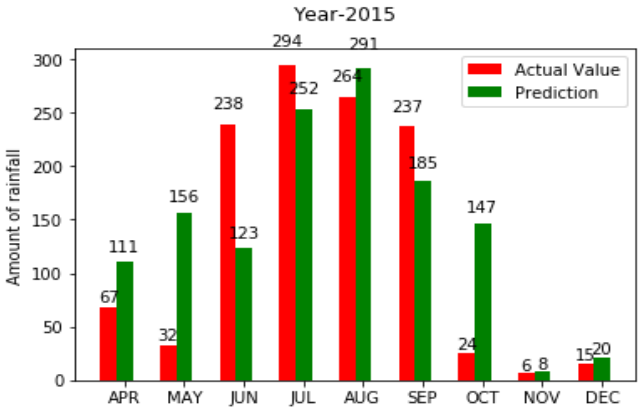

Fig iii. Plot of Actual vs Predicted values of Rainfall in Odisha 2015 for the Odisha data only

B. District wise Prediction Observation:

a. Training on complete dataset

Algorithm MAE

- Linear Regression = 57.08862331011236
- SVR = 116.60671510825178
- Artificial neural nets = 44.132588

b. Training on Odisha dataset only

Algorithm MAE

- Linear Regression = 43.922967624
- SVR = 116.09175497
- Artificial neural nets = 34.43588178

c. Plots

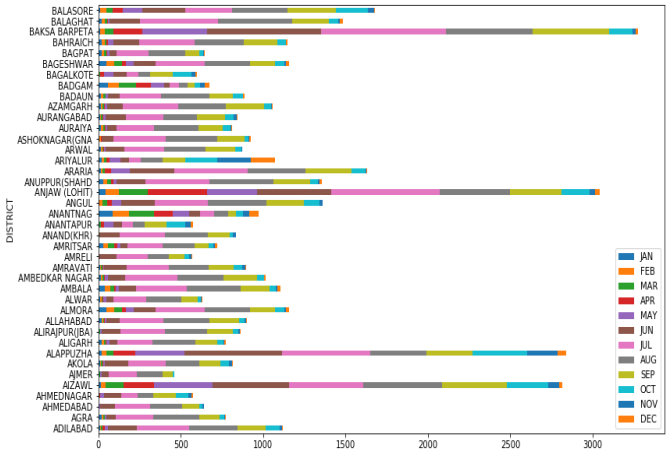

Fig iv. District wise rainfall Odisha in every month part 1

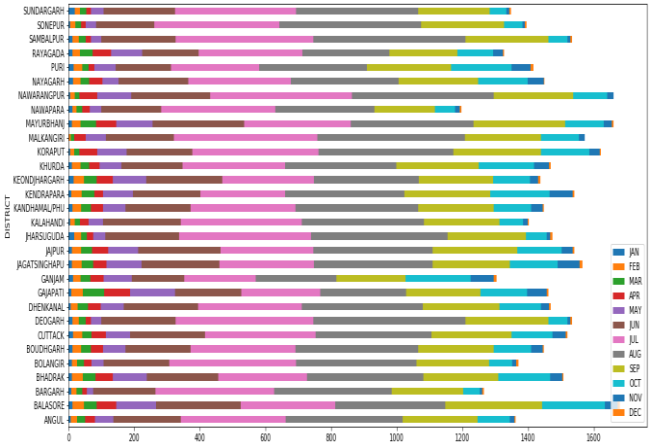

Fig v. District wise rainfall Odisha in every month part 2

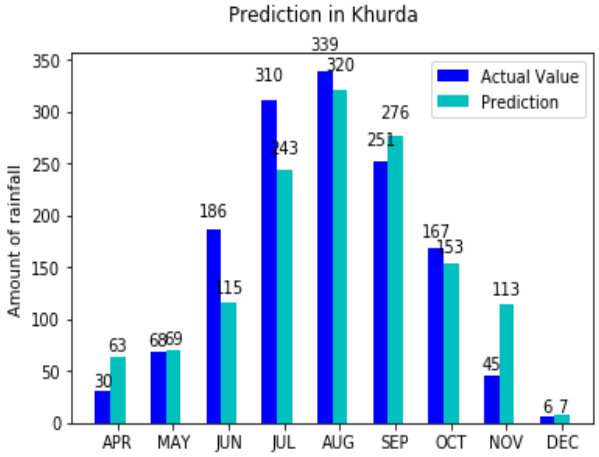

Fig vi. Plot of Actual vs Predicted values of Rainfall In Khurda 2015 for the complete data

6.2. Crop Data Visualization

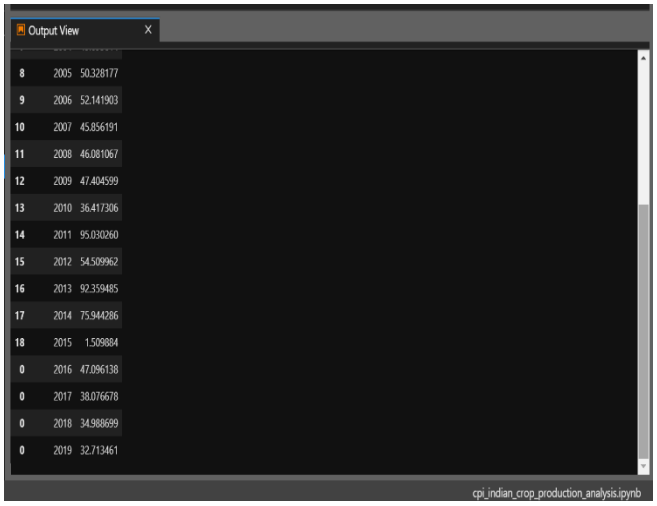

Fig viii. We can see this output screen after the successful compilation of the program. Then we can give inputs to check our desired outputs

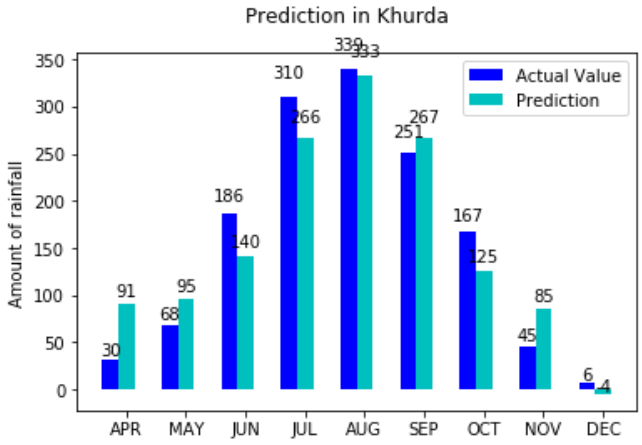

Fig vii. Plot of Actual vs Predicted values of Rainfall in Khurda 2015 for the Odisha data only

Various visualizations of data are observed which helps in implementing the approaches for prediction. Prediction of amount of rainfall for both the types of dataset were observed, and the observations indicate the Neural Networks works much better than any other algorithms for the rainfall dataset.

But however, comparing both the type data I observed that, the MAE value in the district wise data analysis is much lesser than MAE value that was observed for the overall data and the state wise data analysis.

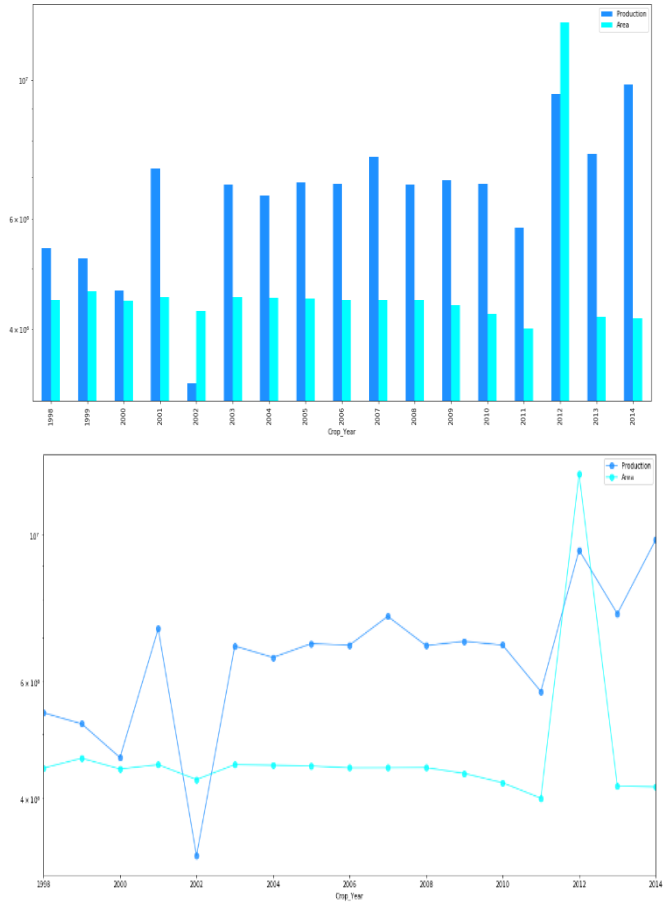

Fig xi. The plot shows the production of rice in Odisha in different years according to the input

6.3. Crop Production Index

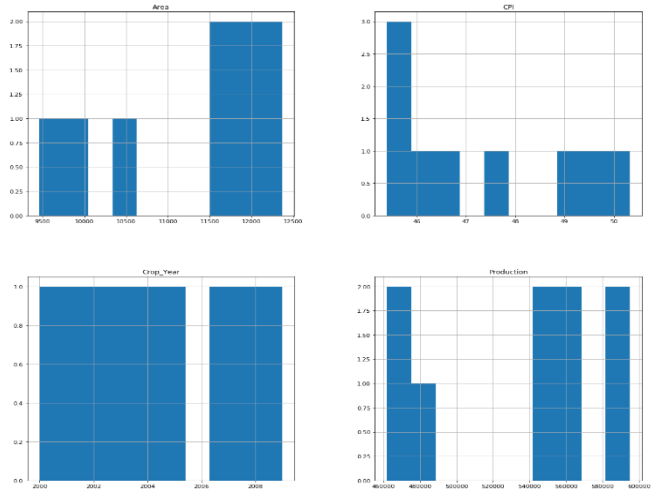

Fig x. Overall analysis of Production, Areas and the CPI for each year

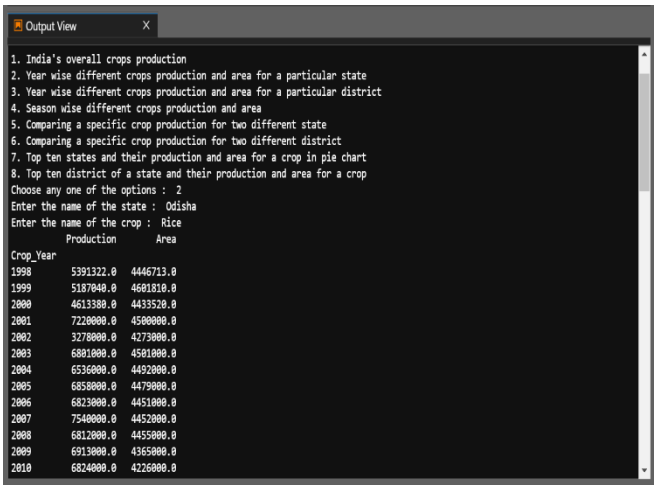

Fig xi. This output shows the predicted values of the CPI and the data from 2016 to 2019 are the forecasted values

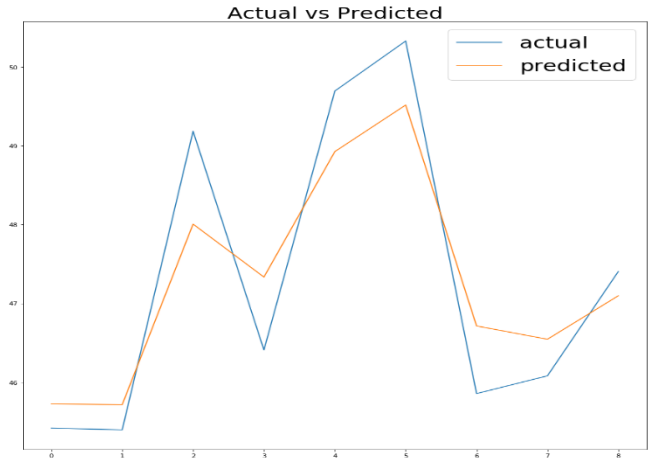

Fig xii. Plot of Actual vs Predicted Values of CPI

The Mean squared error for the program is found out to be 0.53. Hence the technique somehow holds good in predicting the CPI values with least percentage of error.

7. CONCLUSION

The agriculture of a nation depends on several factors and its proper study is immensely useful. In this project, I have tried to get and organize agricultural data in a way in which it can be used for analysis. Visualization helps understanding the data better and hence, the data sets are visualized.

Various visualizations of data are observed which helps in implementing the approaches for prediction.

Prediction of amount of rainfall for both the types of dataset were observed, and the observations indicates the Neural Networks works much better than any other algorithms for the rainfall dataset.

Prediction of the crop production index values are also observed with a mse of 0.53, that mean the program can somehow predict the future values of the crop production index with least accuracy.

REFERENCES

[1] https://data.gov.in/resources/district-rainfall-normal-mm-monthly-seasonal-and-annual-data-period-1951-2000
[2] https://data.gov.in/resources/subdivision-wise-rainfall-and-its-departure-1901-2015
[3] https://data.gov.in/catalog/district-wise-season-wise-crop-production-statistics?filters%5Bfield_catalog_reference%5D=87631&format=json&offset=0&limit=6&sort%5Bcreated%5D=desc
[4] https://analyticsindiamag.com/top-ai-powered-projects-in-indian-agriculture-sector-2019/
[5] http://agriculture.gov.in/
[6] http://agricoop.nic.in/
[7] https://www.indexmundi.com/facts/india/crop-production-index
[8] https://www.researchgate.net/publication/319503839_Machine_Learning_Techniques_For_Rainfall_Prediction_A_Review
[9] https://www.economicsdiscussion.net/agriculture/problems-agriculture/indian-agriculture-problems-7-major-problems-of-indian-agriculture/12859#:~:text=One%20of%20the%20major%20causes,marketing%20presents%20a%20serious%20problem.
[10] https://towardsdatascience.com/predicting-crops-yield-machine-learning-nanodegree-capstone-project-e6ec9349f69
[11] https://iopscience.iop.org/article/10.1088/1748-9326/aae159
[12] http://www.ijesrt.com/issues%20pdf%20file/Archive-2018/April-2018/1.pdf
[13] https://towardsdatascience.com/linear-regression-detailed-view-ea73175f6e86
[14] https://towardsdatascience.com/an-introduction-to-support-vector-regression-svr-a3ebc1672c2
[15] https://medium.com/@rajatgupta310198/getting-started-with-neural-network-for-regression-and-tensorflow-58ad3bd75223
[16] https://www.sigfox.com/en/news/how-rainfall-affects-crop-health
[17] https://www.economicsdiscussion.net/agriculture/problems-agriculture/indian-agriculture-problems-7-major-problems-of-indian-agriculture/12859
[18] https://data.gov.in/

[19] https://knoema.com/atlas/Indonesia/Crop-production-index

BIOGRAPHIES

Premjit Kumar, Undergraduate Student, School of Electronics Engineering, KIIT Deemed to be University, Bhubaneswar is pursuing his Bachelors of Technology (B. Tech) in Electronics and Telecommunication Engineering. He is having skills in various industrial technologies related to Web Technology, Data Science, Machine Learning and Deep Learning. His areas of research are machine learning and deep learning.