Journal DOI: 10.36713/epra1013|SJIF Impact Factor (2021): 7.473 ISSN: 2347-4378
EPRA International Journal of Economics, Business and Management Studies (EBMS)
Volume: 8 | Issue: 6| June 2021 -Peer-reviewed Journal

# SEASONAL CROP DATA ANALYSIS USING PYTHON

**P.Ruth Rini Surajitha**
*Assistant Professor, Department of Management Studies, Centurion University, Vizianagaram-535003*

## ABSTRACT

***Agriculture is the backbone of the Indian economy, but the industry currently needs more support than any other. India is a country of over a billion people in the population, out of which, over 70% of the population lives in the rural areas. With 40% of the country's workforce, agriculture is a major industry and an influence of the Indian economy. Despite this, its contribution to the $2.3 trillion economy is just a meager 16% of the entire GDP and the growth of agriculture sector is about 3.33% only, whereas the growth of our economy is 7% approx. and agriculture contributes only 20% of total exports. The income of farmers is going down and nowadays the youngsters are not showing interest in agriculture for this.***

***Agriculture is one sector responsible for feeding every individual, but the people involved in it are the last to be taken care of. After failing institutions, the time has indeed come for technology to take over the change. With newer problems cropping up every day in the most inevitable indigenous sectors, it is high time we resort to emerging technologies for solutions.***

## OBJECTIVES

**Prediction: -** We are predicting the yield of farmers and our aim is to maximize the income of farmers.
**Identification:-** We are identifying which variables (out of many) have a significant impact on yield.
**Optimization:-** Optimum use of limited resources (Agricultural lands) and get more yield by using our techniques.
**Decision Support Systems for farmers:-** Our aim is to help the farmers in taking their decision regarding the inputs which combination will give maximum output.
**Develop products for farmers: -**Develop new products (Apps, Machinery etc.) according to the requirement.

## SCOPE

**1. Receiving Useful Data to Help Fight Food Scarcity and Empower Small Farmers;-**

Since data scientists have tools to process and analyze gigantic amounts of data efficiently, projects are underway to determine how that information might help small-scale farmers join in the battle to solve income related problem

In December 2018, we started our project, that looks at data from approximately 5000 farmers in impoverished areas from 2 districts of Andhra Pradesh.

The people behind the project hope the data will show whether agricultural investments in various areas are paying off and help develop policies for the farmers. On a larger scale, this project aligns with Andhra Pradesh Sustainable Development Goals to double the agricultural productivity and incomes of farmers in developing nations and help them reduce world hunger.

**2. Managing Crop Diseases and Pests**

Agricultural pests can quickly cut into a farmer's profits. But, misusing pesticides can have adverse effects on people, plants and other living things. Fortunately, we recruit data scientists to help them develop user-facing platforms that analyze when to apply pesticides and how much to use.

Journal DOI: 10.36713/epra1013|SJIF Impact Factor (2021): 7.473 ISSN: 2347-4378
EPRA International Journal of Economics, Business and Management Studies (EBMS)
Volume: 8 | Issue: 6| June 2021 -Peer-reviewed Journal

Our technology relies on Internet of Things (IoT) sensors and artificial intelligence to determine the kind of insects on a crop and the quantity present. Farmers then get an associated report and can use it to plan their pest management approaches. The goal is to help farmers cost-effectively control pests with a minimized environmental impact

**3. Image Recognition**

Employing drones that use artificial intelligence can greatly help farmers to scan their fields, and monitor every stage of the production cycle. This will help farmers to make data-driven decisions. The perfect combination of drone data, I.O.T and other computer technologies can help farmers to develop more sophisticated analysis, and take the agricultural industry to another level making it much more efficient. In fact, latest reports suggest that the use of drones in the agricultural industry is expected to massively rise over the next decade.

**4. Automated Workforce**

According to the World Urbanization Prospects report, 66% of the global population would be in urban areas by the end of 2050. This will lead to a decrease in workforce in the rural areas where the agricultural industry majorly exists. Adopting the latest technologies is the best way to address this situation. Using automated processes empowered by artificial intelligence, many agricultural operations can be remotely done which will help farmers to make faster and smarter decision

**5. Improved Return on Investment**

By including information such as weather conditions, type of soil, marketplace, potential infestations, and data in the algorithm, artificial intelligence can determine the best seed to use and help farmers to maximize the production. This can improve the return on investment for all farms. Further, the AI technology can process analytics that helps farmers to minimize losses in the production supply chain of their farms.

**5. Chat bots**

The agricultural industry can use chat bots as a part of smart technology adoption for providing virtual assistants to farmers. Techniques such as machine learning and deep learning can create chat bots that will answer all questions of the farmers instantly as well as give recommendations and advice over specific problems.While it is true that the artificial intelligence technology offers innumerable benefits to the agricultural industry, it is extremely important to opt for **artificial intelligence services** only from the leading **artificial intelligence service providers** in India having globally recognized certified consultants who have rich expertise in empowering businesses to digitally transform through effective data analytics.

## INPUTS

- Soil Parameters
- Extend of land
- Seed Variety
- Seed Quantity
- Dose of FYM
- Weeding
- Watering
- Fertilizer
- Pesticides

## METHODOLOGY

- **Data collection**
- **Data Entry**
- **Data Validation**
- **Data input**
- **Data Processing**
- **Data analysis**
- **Output**
- **Decision Making**

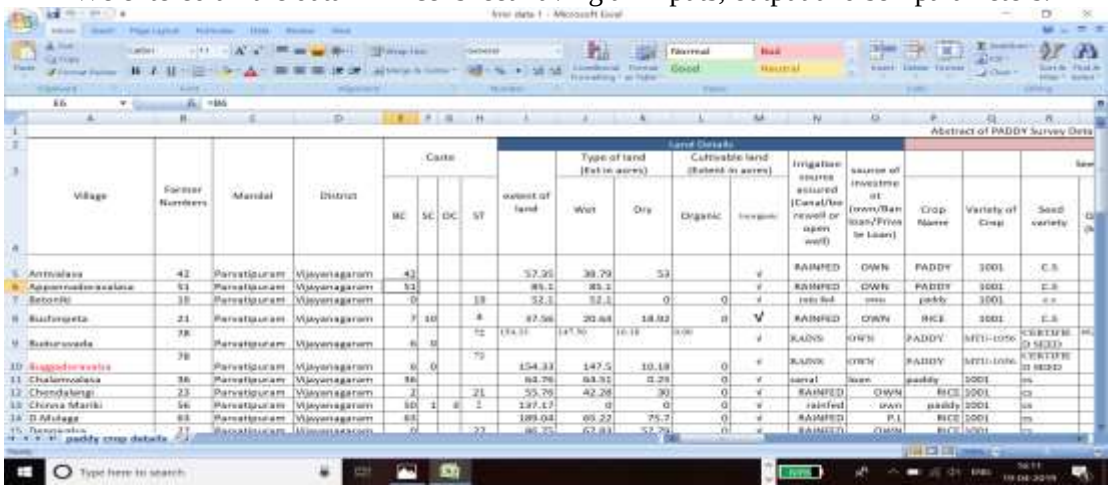
Journal DOI: 10.36713/epra1013|SJIF Impact Factor (2021): 7.473 ISSN: 2347-4378
EPRA International Journal of Economics, Business and Management Studies (EBMS)
Volume: 8 | Issue: 6| June 2021 -Peer-reviewed Journal

## Data Collection

- We collected 5000 farmers' data from Vizianagaram and Srikakulam.
- In this process we have adopted personal interview method of primary data collection.

## DATA ENTRY

We entered all the data in Excel sheet having all inputs, output and soil parameters.



## Data Validation

- In this process we check the data whether the data are entered correctly or not, where there is any blank or not.
- Data validation often referred to as pre-processing is the stage at which row data is cleaned of and organized for the following stage of data processing.
- During data validation the raw data diligently checked for any errors. The purpose of this step is to eliminate bad data (redundant incomplete or incorrect data) and begin to create high quality data for the best *business intelligence.*

*NP-business intelligence means the collection of process, technologies, skills and applications used to make informed, data driven business decision. It facilitated the user in decision making.*

## Data Inputs

- We took our data into "ANACONDA"-(Python platform) as our input.
- The clean data is then entered into its destination and translated into a language that it can be understandable.

The data input is the first stage in which raw data begins to take the form of usable information

Journal DOI: 10.36713/epra1013|SJIF Impact Factor (2021): 7.473 ISSN: 2347-4378
EPRA International Journal of Economics, Business and Management Studies (EBMS)
Volume: 8 | Issue: 6| June 2021 -Peer-reviewed Journal

**Data Processing**

- In this stage we started process the data for interpretation. Processing is done using machine learning algorithm though the process itself may vary slightly depending on the source of data being processed.

**NOTE:-** machine learning is a computer programming technique that uses statistical probabilities to give computers the ability to "learn" without being explicitly programmed.

**Data Analysis**

- In this stage we started define our questions and decide what to measure and decide how to measure it.
- It begins by manipulating our data in a number of deferent ways, such as plotting it out and finding correlations or by creating a pivot table in excel.
- A pivot table let us sort and filter data by different variables and let us calculate the mean, maximum, minimum and standard deviation of our data.
- For the data analysis purpose we use "python" programming language. The output will be discussed in following slides.

NOTE : - python is an interpreted high-level general purpose programming language.

**Output**

- Output or interpretation stage is the stage at which data is finally usable to non technical person.
- It is translated readable and often in form of graphs, images, plain text etc. which enables a person in "decision support system".

NP:- ***decision support system is an information system that supports business organization or a individual in decision making activities.***

- Our analysis is given bellow ->

**P-Value**

- P-value is referred as probability value.
- When we perform a hypothesis test in statistics, a p-value helps us determine the significance of your results. **Hypothesis tests** are used to test the validity of a claim that is made about a population. This claim that's on trial, in essence, is called the **null hypothesis**.
- Here we have taken 0.5 as our p-value to draw our conclusion.
- So here if P-Value will less than 0.5 then that result will statistically significant & vice-versa.

Journal DOI: 10.36713/epra1013|SJIF Impact Factor (2021): 7.473 ISSN: 2347-4378
EPRA International Journal of Economics, Business and Management Studies (EBMS)
Volume: 8 | Issue: 6| June 2021 -Peer-reviewed Journal

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Intercept | 135.1584 | 61.551 | 2.196 | 0.031 | 12.459 | 257.858 |
| EOL | 0.0008 | 6.36e-05 | 13.065 | 0.000 | 0.001 | 0.001 |
| WET_DRY | 13.5150 | 16.070 | 0.841 | 0.403 | -18.520 | 45.550 |
| DRY | 48.1846 | 28.715 | 1.678 | 0.098 | -9.058 | 105.427 |
| WET | -15.3840 | 14.894 | -1.033 | 0.305 | -45.074 | 14.306 |
| DOSEFYM | 1.2464 | 0.095 | 13.065 | 0.000 | 1.056 | 1.437 |
| FERTUREA | -0.3292 | 0.433 | -0.759 | 0.450 | -1.193 | 0.535 |
| FERTDAP | -0.3292 | 0.433 | -0.759 | 0.450 | -1.193 | 0.535 |
| FERTNPKSSP | -0.1646 | 0.217 | -0.759 | 0.450 | -0.597 | 0.267 |
| PDQTY | 119.2524 | 142.036 | 0.835 | 0.407 | -165.486 | 403.990 |
| P2O5 | -0.6811 | 0.236 | -2.890 | 0.005 | -1.151 | -0.211 |
| EC | -1.8329 | 2.047 | -0.504 | 0.615 | -5.115 | 3.049 |
| PH | -7.1153 | 6.262 | -1.136 | 0.260 | -19.599 | 5.369 |
| K | -0.1714 | 0.045 | -3.776 | 0.000 | -0.262 | -0.081 |
| Zn | 18.1000 | 18.505 | 0.546 | 0.587 | -26.709 | 46.909 |
| Fe | -0.0510 | 1.055 | -0.049 | 0.961 | -2.175 | 2.072 |
| Cu | 0.6954 | 0.820 | 0.848 | 0.399 | -0.939 | 2.329 |
| Mn | 0.9529 | 1.344 | 0.709 | 0.481 | -1.727 | 3.632 |
| B | -3.6810 | 3.321 | -1.108 | 0.271 | -10.302 | 2.940 |

P VALUE

| | coef | std err | t | P>\|t\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| Intercept | 70.5328 | 20.517 | 3.438 | 0.001 | 29.733 | 111.332 |
| EOL | 0.0009 | 9.85e-06 | 89.592 | 0.000 | 0.001 | 0.001 |
| DOSEFYM | 1.3239 | 0.015 | 89.592 | 0.000 | 1.295 | 1.353 |
| P2O5 | -0.5626 | 0.102 | -5.533 | 0.000 | -0.765 | -0.360 |
| K | -0.1463 | 0.035 | -4.236 | 0.000 | -0.215 | -0.078 |

SIGNIFICANT

**P VALUE HERE WE HAVE REMOVED THE DATA WHOSE P VALUE IS MORE THAN 0.5**

**R square ($r^2$) :**

R-squared ($R^2$) is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model. Whereas correlation explains the strength of the relationship between an independent and dependent variable, R-squared explains to what extent the variance of one variable explains the variance of the second variable. So, if the $R^2$ of a model is 0.50, then approximately half of the observed variation can be explained by the model's inputs.

Journal DOI: 10.36713/epra1013|SJIF Impact Factor (2021): 7.473 ISSN: 2347-4378
EPRA International Journal of Economics, Business and Management Studies (EBMS)
Volume: 8 | Issue: 6| June 2021 -Peer-reviewed Journal

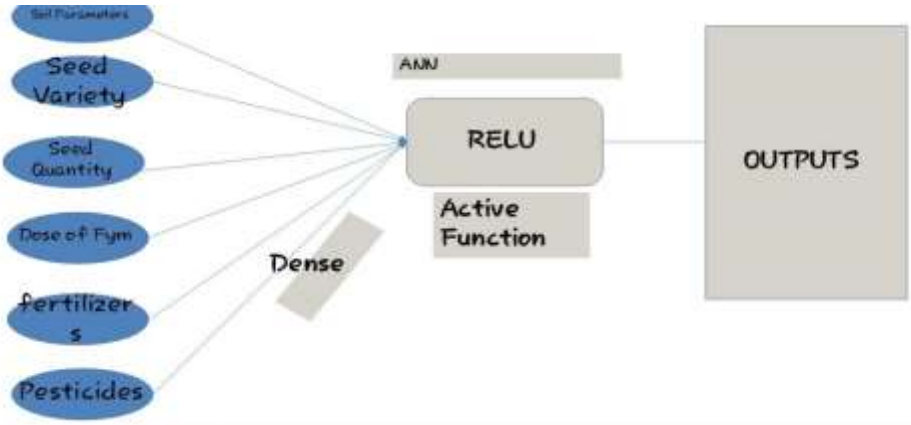


R square value

**TRAIN AND TEST MODEL**

In this process we have to train some part of our data set then use the result model with rest of the data sets

Journal DOI: 10.36713/epra1013|SJIF Impact Factor (2021): 7.473 ISSN: 2347-4378
EPRA International Journal of Economics, Business and Management Studies (EBMS)
Volume: 8 | Issue: 6| June 2021 -Peer-reviewed Journal




## TRAIN AND TEST ANALYSIS

### Mean squared error

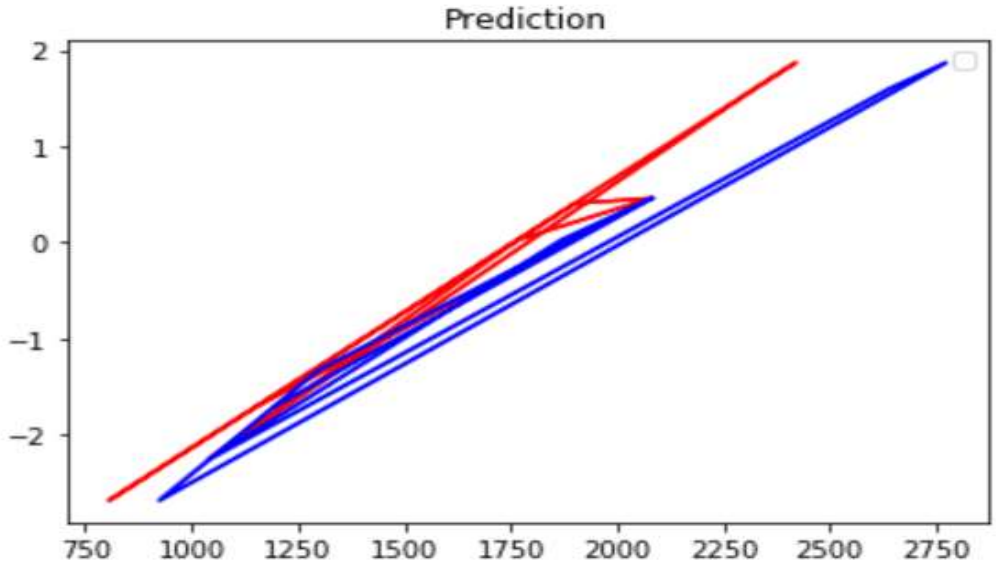
```
In [41]: mean_squared_error(y_test,y_pred)
Out[41]: 4155.643635289521

In [42]: y_test.max()
Out[42]: OPYXG    2496.8
         dtype: float64

In [43]: y_test.min()
Out[43]: OPYXG    886.4
         dtype: float64

In [44]: y_pred.max()
Out[44]: 2597.495255179113

In [45]: y_pred.min()
Out[45]: 880.2846687383156
```

## NEURAL NETWORK

Neural Networks are a class of models within the general machine learning literature. Neural networks are a specific set of algorithms that have revolutionized machine learning. They are inspired by biological neural networks and the current so-called deep neural networks have proven to work quite well.

```
In [6]: model = Sequential()

        # Adding the input layer and the first hidden layer
        model.add(Dense(32, activation = 'relu', input_dim = 2))

        # Adding the second hidden layer
        model.add(Dense(units = 32, activation = 'relu'))

        # Adding the third hidden layer
        model.add(Dense(units = 32, activation = 'relu'))

        # Adding the output layer

        model.add(Dense(units = 1))

        model.compile(loss='mean_absolute_percentage_error', optimizer='adam')

        # Fitting the ANN to the Training set
        model.fit(X_train, y_train, batch_size = 10, epochs = 1000)

        y_pred = model.predict(X_test)
```

Journal DOI: 10.36713/epra1013|SJIF Impact Factor (2021): 7.473 ISSN: 2347-4378
EPRA International Journal of Economics, Business and Management Studies (EBMS)
Volume: 8 | Issue: 6| June 2021 -Peer-reviewed Journal


## CONCLUSION

After the above analysis we found which inputs have significant impact and which combination of variables giving more output.

## SUMMARY

- We use interview method for data collection
- For data inputs we have used excel.
- We have taken help of Python programming for our analysis.
- We got that Dose of Fym, fertilizers & pesticides have an significant impact

## REFERENCES

1. *Introduction to Machine Learning with Python: A Guide for Data Scientists* ***Authors:*** *Andreas C. Muller, Sarah Guido*
2. *Neural Networks and Deep Learning: A Textbook; Book by Charu C. Aggarwal*
3. *Python Machine Learning; Book by Sebastian Raschka*
4. *Neural Networks and Learning Machines, 3d Edition; Book by Simon S. Haykin*
5. *Artificial Intelligence: A Modern Approach; Textbook by Peter Norvig and Stuart J. Russell*
6. *Excel 2020: Learn Excel Essential Skills with Smart Method; Book by Alex PARKER*