

INTERNSHIP REPORT ON

“A PREDICTIVE MODEL FOR FORECASTING DEMAND AND SUPPLY INFORMATION OF TOP CROPS”

Submitted in partial fulfilment for the award of degree
BACHELOR OF ENGINEERING
in
COMPUTER SCIENCE AND ENGINEERING

Submitted by:
NAME: SHAMSHUDDIN ABDUL MUBIN KHAN
USN: 2GJ19CS006



Conducted at
COMPSOFT TECHNOLOGIES



Department of Computer Science and Engineering

GIRIJABAI SAIL INSTITUTE OF TECHNOLOGY, KARWAR
Utkagali, Adjacent to Karnataka-Goa Border check post,
NH 66, Majali, Karwar, Uttara Kannada - 581345

GIRIJABAI SAIL INSTITUTE OF TECHNOLOGY, KARWAR
Utkagali, Adjacent to Karnataka-Goa Border check post,
NH 66, Majali, Karwar, Uttara Kannada - 581345



CERTIFICATE

This is to certify that the Internship titled “**A Predictive Model For Forecasting Demand And Supply Information Of Top Crops**” carried out by **Shamshuddin Abdul Mubin Khan**, a bonafide student of **Girijabai Sail Institute Of Technology**, in partial fulfillment for the award of **Bachelor of Engineering**, in **Computer Science** under Visvesvaraya Technological University, Belagavi, during the year 2022-2023. It is certified that all corrections/suggestions indicated have been incorporated in the report.

Signature of Guide

Signature of HOD

Signature of Principal

External Viva:

Name of the Examiner

Signature with Date

1)_____

2)_____

D E C L A R A T I O N

I, Shamshuddin Abdul Mubin Khan, final year student of Computer Science, Girijabai Sail Institute of Technology, declare that the Internship has been successfully completed in Compsoft Technologies. This report is submitted in partial fulfillment of the requirements for award of Bachelor Degree in Computer Science, during the academic year 2022-2023.

Date :

Place :

USN : 2GJ19CS006

NAME : SHAMSHUDDIN ABDUL MUBIN KHAN



Internship Offer Letter

Date: **23rd August, 2022**

Name: **Shamshuddin Abdul Mubin Khan**

USN: **2GJ19CS006**

Dear Student,

We would like to congratulate you on being selected for the **Machine Learning With Python(Research Based)** Internship position with **Compsoft Technologies**, effective Start Date **23rd August, 2022**, All of us are excited about this opportunity provided to you!

This internship is viewed as being an educational opportunity for you, rather than a part-time job. As such, your internship will include training/orientation and focus primarily on learning and developing new skills and gaining a deeper understanding of concepts of **Machine Learning With Python(Research Based)** through hands-on application of the knowledge you learn while you train with the senior developers. You will be bound to follow the rules and regulations of the company during your internship duration.

Again, congratulations and we look forward to working with you!.

Sincerely,

Nithin K. S
Project Manager
Compsoft Technologies
*No. 363, 19th main road,
1st Block Rajajinagar
Bangalore - 560010*

ACKNOWLEDGEMENT

I want to thank my advisers and everyone at the company for their patience and assistance during my on-site training. Thanks to their guidance, I was able to develop and learn about the **Machine Learning with Python**. These skills will help me to expand my resume and advance my career.

I undertook this internship project and completed the internship report under the guide and student coordinator and Compsoft Technologies. I'm grateful to the Compsoft staff for their patience and assistance during my training at their company.

USN : 2GJ19CS006

NAME : SHAMSHUDDIN ABDUL MUBIN KHAN

ABSTRACT

Agriculture is the biggest industry in India and it generates a significant number of employments in the country. The features of weather, geography, and soil of India are diverse. As a consequence, a range of crops are grown in the country. India crops include food grains like rice, wheat, and pulses.

The increase in population will be more in developing countries like India. When the price of any commodities set too high then the suppliers try to produce more goods to make more profit. Conversely, if the supply is less for any commodities, as consumers have to compete with one other to buy the less supplied goods, results in increased price for the commodity, making consumers suffer with the high price.

As there is no synchronization in production and demand for the agricultural commodities, either farmer fail to get good market prices for their products, or consumer suffers high prices due to less production.

Around 42% of the people depend on agriculture for their livelihood. The economic upliftment of farmers happens when there is a seamless transfer of agricultural produce from producers to the consumers. It is evident that there is a huge gap between demand and supply of various crops, due to which both farmers and consumers are facing problems. At present, in India there is no system in place to efficiently manage this demand and supply issue. The potential of present-day technologies like data analytics, machine learning can be exploited to overcome these issues. The available data about the demand, supply, price variation of the crops and other factors affecting the supply chain of agricultural produce can be used to analyse and come up with a model to predict and forecast market variations of agricultural crops.

TABLE OF CONTENTS

Sl no	Description	Page no
1	Learning Objectives/Internship Objectives	8-8
2	Company Profile	8-9
3	About the Company	9-12
4	Introduction	12-13
5	System Analysis	13-14
6	. Software Requirement Specifications	14-14
7	Design Analysis	15-16
8	Implementation	16-17
9	Coding	17-20
10	Snapshots	20-27
11	Conclusion	29
12	References	30

1. Learning Objectives/Internship Objectives

- Internships are generally thought of to be reserved for college students looking to gain experience in a particular field. However, a wide array of people can benefit from Training Internships in order to receive real world experience and develop their skills.
- An objective for this position should emphasize the skills you already possess in the area and your interest in learning more
- Internships are utilized in a number of different career fields, including architecture, engineering, healthcare, economics, advertising and many more.
- Some internship is used to allow individuals to perform scientific research while others are specifically designed to allow people to gain first-hand experience working.
- Utilizing internships is a great way to build your resume and develop skills that can be emphasized in your resume for future jobs. When you are applying for a Training Internship, make sure to highlight any special skills or talents that can make you stand apart from the rest of the applicants so that you have an improved chance of landing the position.

2. COMPANY PROFILE

A Brief History of Compsoft Technologies

Compsoft Technologies, was incorporated with a goal "To provide high quality and optimal Technological Solutions to business requirements of our clients". Every business is a different and has a unique business model and so are the technological requirements. They understand this and hence the solutions provided to these requirements are different as well. They focus on clients requirements and provide them with tailor made technological solutions. They also understand that Reach of their Product to its targeted market or the automation of the existing process into e-client and simple process are the key features that our clients desire from Technological Solution they are looking for and these are the features that we focus on while designing the solutions for their clients.

Sarvamoola Software Services. is a Technology Organization providing solutions for all web design and development, MYSQL, PYTHON Programming, HTML, CSS, ASP.NET and

LINQ. Meeting the ever increasing automation requirements, Sarvamoola Software Services. specialize in ERP, Connectivity, SEO Services, Conference Management, effective web promotion and tailor-made software products, designing solutions best suiting clients requirements.

Compsoft Technologies, strive to be the front runner in creativity and innovation in software development through their well-researched expertise and establish it as an out of the box software development company in Bangalore, India. As a software development company, they translate this software development expertise into value for their customers through their professional solutions.

They understand that the best desired output can be achieved only by understanding the clients demand better. Compsoft Technologies work with their clients and help them to define their exact solution requirement. Sometimes even they wonder that they have completely redefined their solution or new application requirement during the brainstorming session, and here they position themselves as an IT solutions consulting group comprising of high caliber consultants.

They believe that Technology when used properly can help any business to scale and achieve new heights of success. It helps Improve its efficiency, profitability, reliability; to put it in one sentence " Technology helps you to Delight your Customers" and that is what we want to achieve.

3. ABOUT THE COMPANY



Compsoft Technologies is a Technology Organization providing solutions for all web design and development, MYSQL, PYTHON Programming, HTML, CSS, ASP.NET and LINQ. Meeting the ever increasing automation requirements, Compsoft Technologies specialize in ERP, Connectivity, SEO Services, Conference Management, effective web promotion and tailor-made software products, designing solutions best suiting clients requirements. The organization where they have a right mix of professionals as stakeholders to help us serve our clients with best of our capability and with at par industry standards. They have young, enthusiastic, passionate and creative Professionals to develop technological innovations in the field of Mobile technologies, Web applications as well as Business and Enterprise solution. Motto of our organization is to "Collaborate with our clients to provide them with best Technological solution hence creating Good Present and Better Future for our client which will

bring a cascading a positive effect in their business shape as well”. Providing a Complete suite of technical solutions is not just our tag line, it is Our Vision for Our Clients and for Us, We strive hard to achieve it.

Products of Compsoft Technologies.

Android Apps

It is the process by which new applications are created for devices running the Android operating system. Applications are usually developed in Java (and/or Kotlin; or other such option) programming language using the Android software development kit (SDK), but other development environments are also available, some such as Kotlin support the exact same Android APIs (and bytecode), while others such as Go have restricted API access.

The Android software development kit includes a comprehensive set of development tools. These include a debugger, libraries, a handset emulator based on QEMU, documentation, sample code, and zutorials. Currently supported development platforms include computers running Linux (any modern desktop Linux distribution), Mac OS X 10.5.8 or later, and Windows 7 or later. As of March 2015, the SDK is not available on Android itself, but softwaredevelopment is possible by using specialized Android applications.

Web Application

It is a client–server computer program in which the client (including the user interface and client-side logic) runs in a web browser. Common web applications include web mail, online retail sales, online auctions, wikis, instant messaging services and many other functions. web applications use web documents written in a standard format such as HTML and JavaScript, which are supported by a variety of web browsers. Web applications can be considered as a specific variant of client–server software where the client software is downloaded to the client machine when visiting the relevant web page, using standard procedures such as HTTP. The Client web software updates may happen each time the web page is visited. During the session, the web browser interprets and displays the pages, and acts as the universal client for any web application. The use of web application frameworks can often reduce the number of errors in a program, both by making the code simpler, and by allowing one team to concentrate on the framework while another focuses on a specified use case. In applications which are exposed to constant hacking attempts on the Internet, security-related problems can be caused by errors in the program.

Frameworks can also promote the use of best practices such as GET after POST. There are some who view a web application as a two-tier architecture. This can be a “smart” client that performs all the work and queries a “dumb” server, or a “dumb” client that relies on a “smart” server. The client would handle the presentation tier, the server would have the database (storage tier), and the business logic (application tier) would be on one of them or on both.

While this increases the scalability of the applications and separates the display and the

database, it still doesn't allow for true specialization of layers, so most applications will outgrow this model. An emerging strategy for application software companies is to provide web access to software previously distributed as local applications. Depending on the type of application, it may require the development of an entirely different browser-based interface, or merely adapting an existing application to use different presentation technology. These programs allow the user to pay a monthly or yearly fee for use of a software application without having to install it on a local hard drive. A company which follows this strategy is known as an application service provider (ASP), and ASPs are currently receiving much attention in the software industry.

Security breaches on these kinds of applications are a major concern because it can involve both enterprise information and private customer data. Protecting these assets is an important part of any web application and there are some key operational areas that must be included in the development process. This includes processes for authentication, authorization, asset handling, input, and logging and auditing. Building security into the applications from the beginning can be more effective and less disruptive in the long run.

Web design

It encompasses many different skills and disciplines in the production and maintenance of websites. The different areas of web design include web graphic design; interface design; authoring, including standardized code and proprietary software; user experience design; an search engine optimization. The term web design is normally used to describe the design process relating to the front-end (client side) design of a website including writing mark up. Web design partially overlaps web engineering in the broader scope of web development. Web designers are expected to have an awareness of usability and if their role involves creating markup then they are also expected to be up to date with web accessibility guidelines. Web design partially overlaps web engineering in the broader scope of web development.

Departments and services offered

Compsoft Technologies plays an essential role as an institute, the level of education, development of student's skills are based on their trainers. If you do not have a good mentor then you may lag in many things from others and that is why we at Compsoft Technologies gives you the facility of skilled employees so that you do not feel unsecured about the academics. Personality development and academic status are some of those things which lie on mentor's hands. If you are trained well then you can do well in your future and knowing its importance of Compsoft Technologies always tries to give you the best.

They have a great team of skilled mentors who are always ready to direct their trainees in the best possible way they can and to ensure the skills of mentors we held many skill development programs as well so that each and every mentor can develop their own skills with the demands of the companies so that they can prepare a complete packaged trainee.

Services provided by Compsoft Technologies.

- Core Java and Advanced Java
- Web services and development
- Dot Net Framework
- Python
- Selenium Testing
- Conference / Event Management Service
- Academic Project Guidance
- On The Job Training
- Software Training

4. INTRODUCTION

The world's population may reach 6.3 billion and this population growth may surge to 9.5 to 10.0 billion by 2050. The increase in population will be more in developing countries like India than developed countries. The economic growth also has been increased dramatically in the recent years in such countries. The increase in income results in increased demand for more and better food. In India, for instance, the increased income has doubled the expenditure on food commodities. The same pattern has been observed in all the developing countries. It has been calculated that, due to increase in population and individual's income, the world food demand will double by 2050. In topical years, there has been a significant variation in the rates of agricultural products like fruits, flowers, food- grains and vegetables, etc. As shown in Fig 1, the price variation is mainly due to the mismatch in demand and supply of these agricultural products. When the price of any commodity is set too high then the suppliers (farmers) tries to produce more goods to make more profit. When the price of any commodity is set too high then the consumers will tend to purchase less because of high rate, and the suppliers (farmers) incurs the loss. Conversely, if the supply is less for any commodities, as consumers have to compete with one other to buy the less supplied goods, results in increased price for the Hence, in order to reduce the mismatch in demand and supply of food crops effectively, primarily the expected demand for various food commodities needs to be forecasted and guide the farmers accordingly. So there is a need for some system that could guide the farmers in selecting and growing the crops to satisfy the actual demand of the society. This could eliminate the gap between the consumer's demand and producer's supply and reducing the loss for both consumers and farmers. Big data analytics as an emerging trend could help in providing solutions for such problems. An effective forecasting model is proposed and has been implemented in this paper that (i) determine the gap between the demand for and supply of the crops that have to be reduced. (ii) Forecasts the demand of various food commodities that helps the system to guide the farmers in selecting and growing the appropriate crops to satisfy the demand and hence reducing the gap or mismatch between the demand and supply of the crops. For this purpose, the huge data sets have been collected from various sources like government websites and elaborated market survey for the years 2005-2016. Using these data, forecasting model is built for onion, tomato and rice data sets for the state of

Karnataka. Also compared the forecasted values of the same commodities for the year 2017 market data and the results have shown the promising equilibrium.

The remaining portion of the paper is organized as follows in section II, related work has been discussed. In section III the proposed Demand-Prediction Forecasting Model[DPFM] has been described, section IV illustrates the implementation of the model, section V evaluates the results and finally, section VI concludes the work

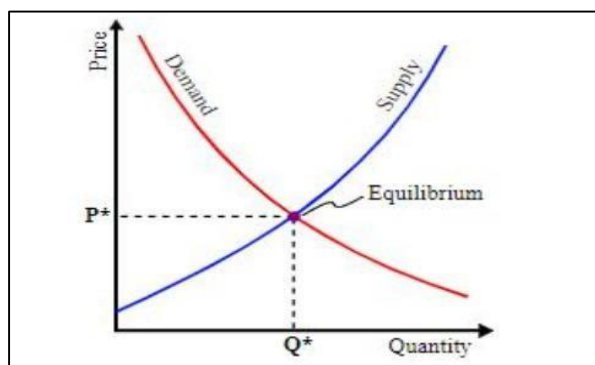


Fig. 1. Demand, supply and price inflation

5. SYSTEM ANALYSIS

Existing System:

The necessary data required for this analysis has been gathered from the sources like Ministry-of-Agriculture, Agmarknet, Directorate of marketing and Inspection, Ministry-of-Agriculture and Farmers-Welfare, Government of India, National Horticulture Board (NHB) India, HOPCOMS Horticultural Producer's Cooperative Marketing and Processing Society Limited by applying Web scraping methods and stored in the local repository and detailed market survey. Data pre-processing module removes the noise from the collected data sets and builds the missing values before applying forecasting algorithm for better performance. All the data sets collected were integrated into a single dataset. During this process, incomplete information is eliminated and all NA (Not Applicable) values are aggregated to the average value. The data sets collected contains multiple attributes, the required attributes are separated and stored as a separate data frame, and then the data frame is converted to time series data. The pre-processed data has been stored and processed as clusters in distributed mode for effective application of the algorithm and analysis. The Hadoop HDFS and Map Reduce paradigm have been used to provide a distributed data storing and parallel

Proposed System:

The proposed DPFM model is implemented using the integrated R-Hadoop machine learning based prediction modeling that provides scalable and parallel processing environment. Also, a Map-Reduce programming model has been developed in R environment to perform the efficient analysis of the data stored in Hadoop clusters. The server that runs R submits the jobs to Hadoop

which in turn distributes the work among m machines in the cluster and gets the result.

6. SOFTWARE REQUIREMENT SPECIFICATIONS

6.1: System configurations

The software requirement specification can produce at the culmination of the analysis task. The function and performance allocated to software as part of system engineering are refined by established a complete information description, a detailed functional description, a representation of system behavior, and indication of performance and design constrain, appropriate validate criteria, and other information pertinent to requirements.

6.1.1: Hardware Requirement:

- System : i3 2.4 GHz & above
- Hard Disk : 256GB
- Ram : 4GB

6.1.2 : Software Requirements:

- Operating system : Windows 10
- Coding Language : Python
- Application : Jupyter Notebook.

. There are a lot of python libraries which could be used to build visualization like *matplotlib*, *vispy*, *bokeh*, *seaborn*, *pygal*, *folium*, *plotly*, *cufflinks*, and *networkx*. Of the many, *matplotlib* and *seaborn* seems to be very widely used for basic to intermediate level of visualizations.

However, two of the above are widely used for visualization i.e.

- **Matplotlib:** It is an amazing visualization library in Python for 2D plots of arrays, It is a multi-platform data visualization library built on *NumPy* arrays and designed to work with the broader *SciPy* stack. Use the below command to install this library:

```
pip install matplotlib
```

- **Seaborn:** This library sits on top of *matplotlib*. In a sense, it has some flavors of *matplotlib* while from the visualization point, its is much better than *matplotlib* and has added features as well. Use the below command to install this library:

```
pip install seaborn
```

6.2: Step-by-step Approach

- Import required modules
- Load the dataset.
- Display the data and constraints of the loaded dataset.
- Use different methods to visualize various illustrations from the data.

7. DESIGN ANALYSIS

This consists of modules for data collection, data repository, data pre-processing, clustering, MapReduce and forecasting the demand.

7.1: SYSTEM ARCHITECTURE

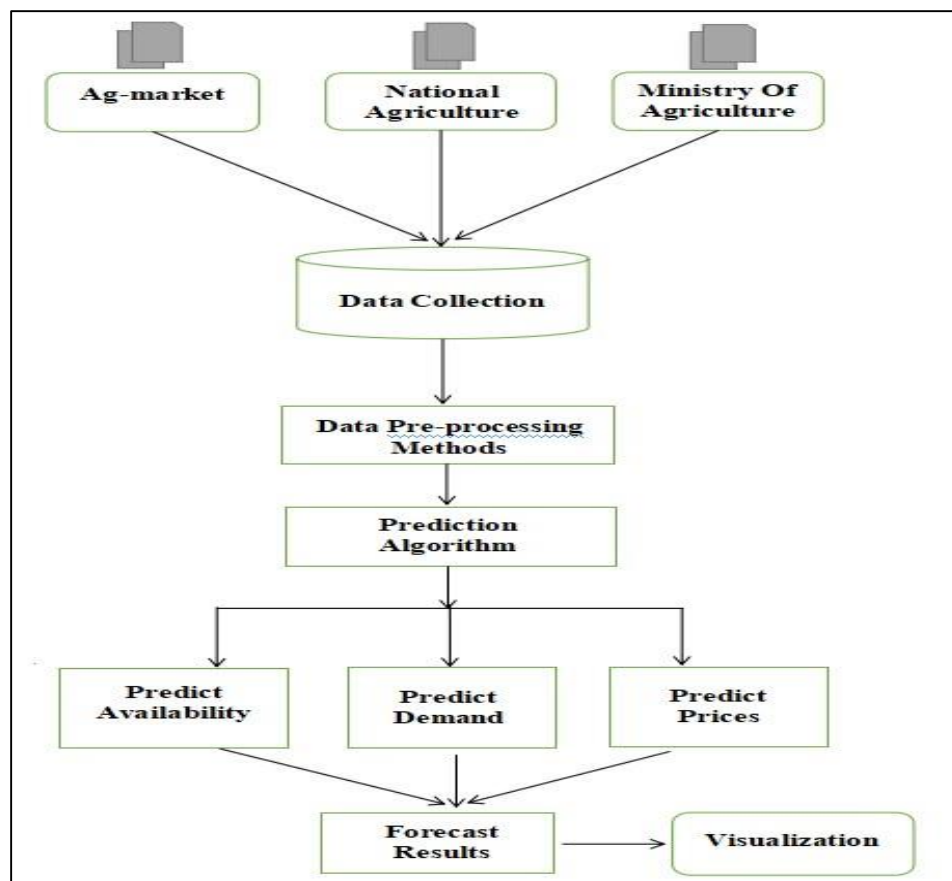


Fig. 2. System Architecture

The process of transporting data from one or more sources to a target site for further processing and analysis. Agricultural data of previous years are collected and used by the system. This dataset includes crop areas, types of crops cultivated, nature of the soil, yields and overall crops consumption. Data is gathered

from authentic websites like Ministry of Agriculture & Farmers Welfare, Food and Agriculture Organization, APEDA, NITI Aayog, Agriculture Marketing Department of Karnataka, Indiatat.com and Competition Commission. Additionally, some unpublished data has also been procured from APMCs and district agriculture and horticulture departments. The weather condition data is collected from authentic sources like IMD.

Crop prices are affected by several factors such as climate, supply and demand. The obtained data contained huge number of outliers, null values and many discontinuous values. An outlier is a data point that is noticeably different from the rest. They represent errors in measurement, bad data collection, or simply show variables not considered when collecting the data. Learning algorithms are sensitive to outliers. Using Python libraries in Excel we managed to reduce the outliers and error values. Since demand data was unaccounted in any authentic websites, simulated data has been used for analysis. We calculated this data by making use of demand curve formula which fits the curve. Yearly data are collected for forecasting because it has less noise. As the data for demand was simulated, the required accuracy is partially met, when actual data is available, accuracy of the model can be increased.

$$m = \frac{\sum_{i=1}^n y_2 - y_1}{x_2 - x_1}$$

y = Max Price (Rs.
/Quintal) x = Production
(Kg)

The equation that depicts the relationship between the price of a certain commodity and the quantity of that commodity that is demanded at that price can be given as

$$Q_d = a - mP$$

Q_d = Linear demand
curve a = Production
(Kg) m= Slope
P = Price (Rs)

The algorithms and tools thus selected are familiarized by carrying out some test runs and finding the most optimal algorithm to satisfy the needs. Algorithms like Linear regression, Logistic regression and Random Forest are used for prediction and classification. Linear regression is initially carried out to predict the value of a variable based on the value of another independent variable. The chosen algorithms are then implemented in sequence to design a predictive model.

Model validation is carried out in two phases. In the initial phase, real time data is given as an input to the designed predictive model to obtain the forecasting information. These results are compared, verified and validated against the authentic data to check for accuracy. In the second phase, Orange3, a python based data visualization, machine learning and data mining tool kit has been used for explorative rapid qualitative data analysis to validate our prediction model

8. IMPLEMENTATION

Implementation is the stage where the theoretical design is turned into a working system. The most crucial stage in achieving a new successful system and in giving confidence on the new system for the users that it will work efficiently and effectively.

The system can be implemented only after thorough testing is done and if it is found to work according to the specification. It involves careful planning, investigation of the current system and its constraints on implementation, design of methods to achieve the change over and an evaluation of change over methods as a part from planning.

Two major tasks of preparing the implementation are education and training of the users and testing of the system. The more complex the system being implemented, the more involved will be the system analysis and design effort required just for implementation.

The implementation phase comprises of several activities. The required hardware and software acquisition is carried out. The system may require some software to be developed. For this, programs are written and tested. The user then changes over to his new fully tested system and the old system is discontinued.

TESTING

The testing phase is an important part of software development. It is the Information zed system will help in automate process of finding errors and missing operations and also a complete verification to determine whether the objectives are met and the user requirements are satisfied. Software testing is carried out in three steps:

1. The first includes unit testing, where in each module is tested to provide its correctness, validity and also determine any missing operations and to verify whether the objectives have been met. Errors are noted down and corrected immediately.
2. Unit testing is the important and major part of the project. So errors are rectified easily in particular module and program clarity is increased. In this project entire system is divided into several modules and is developed individually. So unit testing is conducted to individual modules.
3. The second step includes Integration testing. It need not be the case, the software whose modules when run individually and showing perfect results, will also show perfect results when run as a whole.

9: CODING

Input[1]:

```
# importing pandas module
import pandas as pd

# load the dataset
data = pd.read_csv('crop.csv')

# display top 5 values
data.head()
```

Input[2]:

```
# data description
data.info()
```

Input[3]:

```
# 2011 crop data in histogram analysis
data['2011'].hist()
```

Input[4]:

```
# 2012 crop data in histogram analysis
data['2012'].hist()
```

Input[4]:

```
# 2013 crop data in histogram analysis
data['2013'].hist()
```

Input[5]:

```
# display all year data
data.hist()
```

Input[6]:

```
# import seaborn module

import seaborn as sns

# setting style

sns.set_style("whitegrid")

# plotting data using boxplot for 2013 - 2014

sns.boxplot(x='2013', y='2014', data=data)
```

Input[7]:

```
# scatter plot 2013 data vs 2014 data

plt.scatter(data['2013'],data['2014'])

plt.show()
```

Input[8]:

```
# line plot 2013 data vs 2014 data

plt.plot(data['2013'],data['2014'])

plt.show()
```

Input[9]:

```
# import required modules
import matplotlib.pyplot as plt
from scipy import stats

# assign data
x = data['2017']
y = data['2018']

# linear regression 2017 data vs 2018 data
slope, intercept, r, p, std_err = stats.linregress(x, y)

# function to return slope
def myfunc(x):
    return slope * x + intercept

mymodel = list(map(myfunc, x))

# scatter
plt.scatter(x, y)

# plotting the data
plt.plot(x, mymodel)

# display the figure
plt.show()

# import required modules
import matplotlib.pyplot as plt
from scipy import stats

# assign data
x = data['2016']
y = data['2017']

# linear regression 2017 data vs 2018 data
slope, intercept, r, p, std_err = stats.linregress(x, y)

# function to return slope
def myfunc(x):
    return slope * x + intercept

mymodel = list(map(myfunc, x))

# scatter
plt.scatter(x, y)

# plotting the data
plt.plot(x, mymodel)

# display the figure
plt.show()
```

Input[10]:

```

# import required modules
import matplotlib.pyplot as plt
from scipy import stats

# assign data
x = data['2016']
y = data['2017']

# linear regression 2017 data vs 2018 data
slope, intercept, r, p, std_err = stats.linregress(x, y)

# function to return slope
def myfunc(x):
    return slope * x + intercept
mymodel = list(map(myfunc, x))

# scatter
plt.scatter(x, y)

# plotting the data
plt.plot(x, mymodel)

# display the figure
plt.show()

```

10: SNAPSHOTS

OUTPUT 1:

	Crop	2011	2012	2013	2014	2015	2016	2017	2018
0	Rice	100	101	99	105	112	121	117	110
1	Wheat	100	101	112	115	117	127	120	108
2	Coarse Cereals	100	107	110	115	113	123	122	136
3	Pulses	100	108	134	124	124	146	137	129
4	Vegetables	100	109	103	118	113	124	128	115

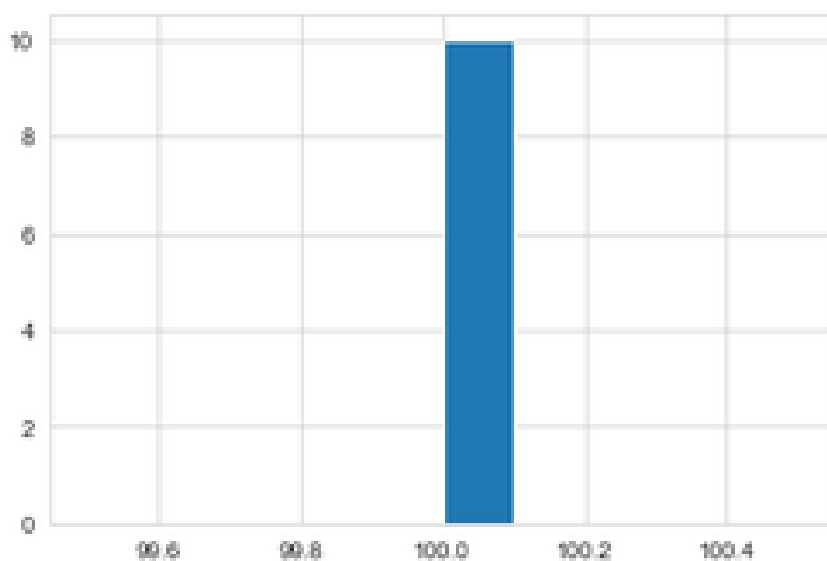
OUTPUT 2:

```
In [11]: data.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10 entries, 0 to 9
Data columns (total 9 columns):
#   Column  Non-Null Count  Dtype  
---  -
0   Crop    10 non-null     object  
1   2011    10 non-null     int64   
2   2012    10 non-null     int64   
3   2013    10 non-null     int64   
4   2014    10 non-null     int64   
5   2015    10 non-null     int64   
6   2016    10 non-null     int64   
7   2017    10 non-null     int64   
8   2018    10 non-null     int64   
dtypes: int64(8), object(1)
memory usage: 848.0+ bytes
```

OUTPUT 3:

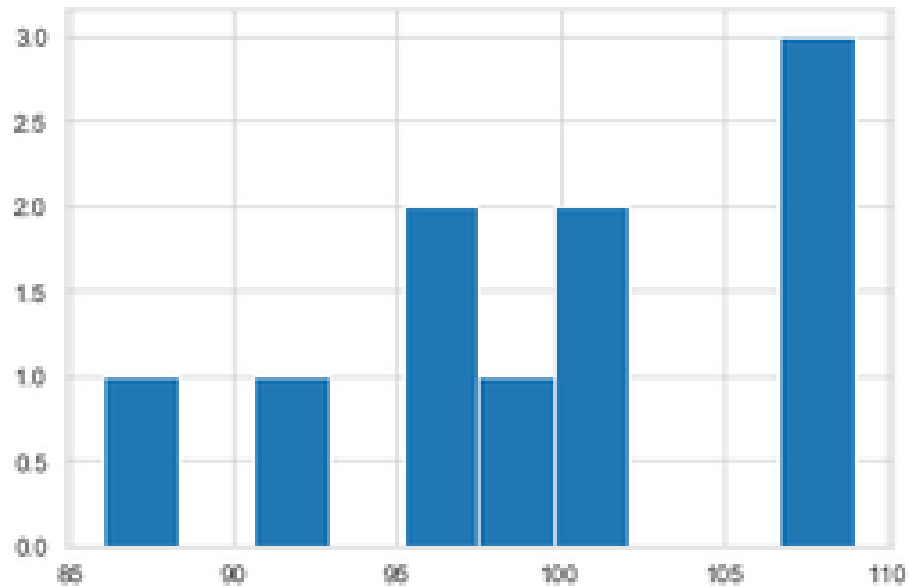
```
<matplotlib.axes._subplots.AxesSubplot at 0x1c556985448>
```



The above program depicts the crop production data in the year 2011 using histogram.

OUTPUT 3:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c5557cd448>
```

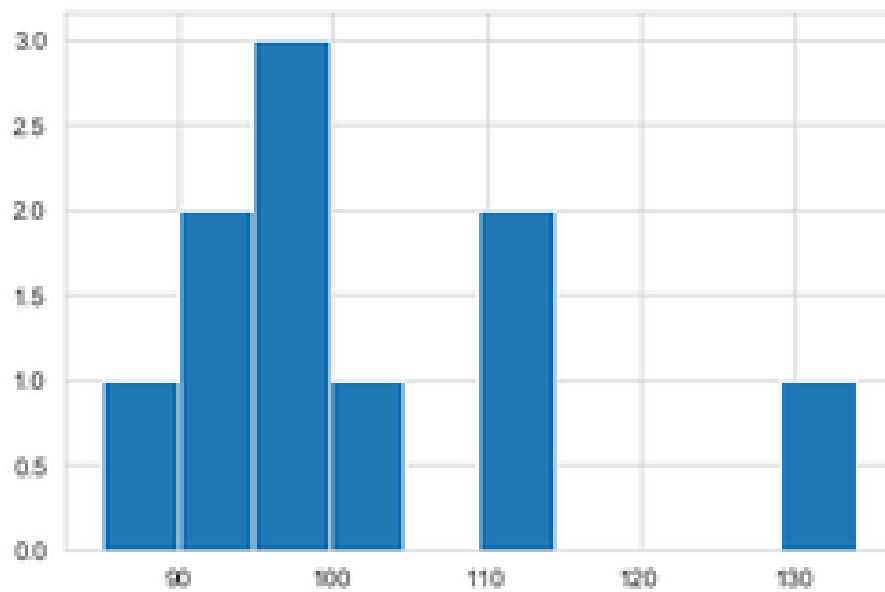


The above program depicts the crop production data in the year 2012 using histogram.

OUTPUT 4:

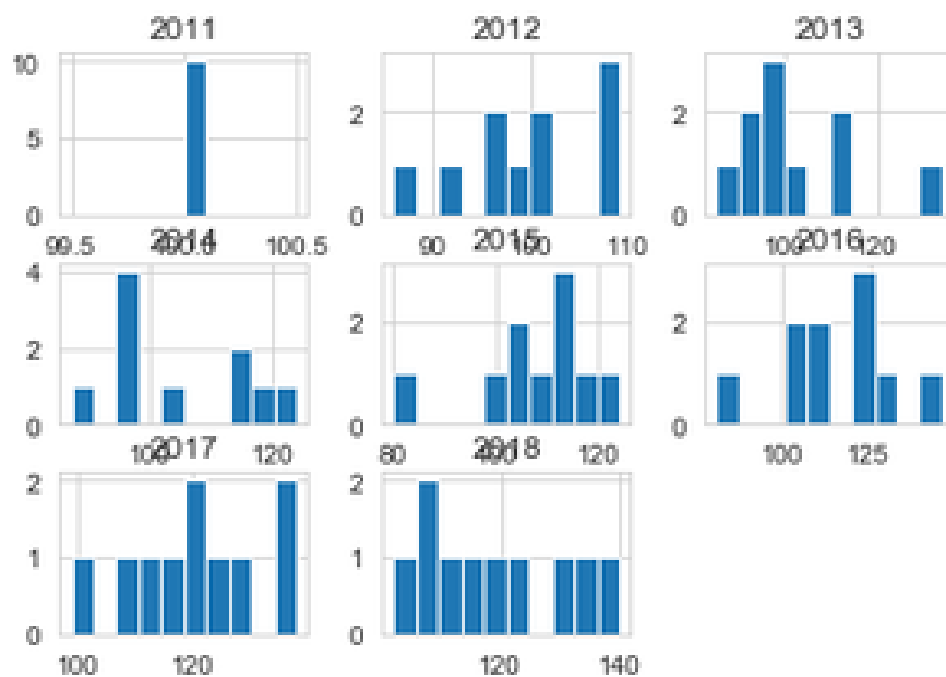
```
data['2013'].hist()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1c55586be48>
```



The above program depicts the crop production data in the year 2013 using histogram.

OUTPUT 5:



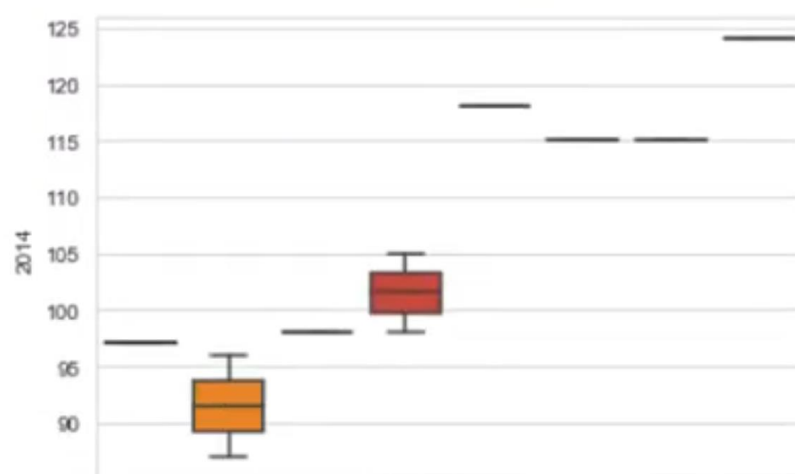
The above program depicts the crop production data of all the available time periods(year) using multiple histograms.

OUTPUT 6:

```
#setting style
```

```
sns.boxplot(x = '2013', y = '2014', data = data)  
#plotting data using boxplot for 2013 - 2014
```

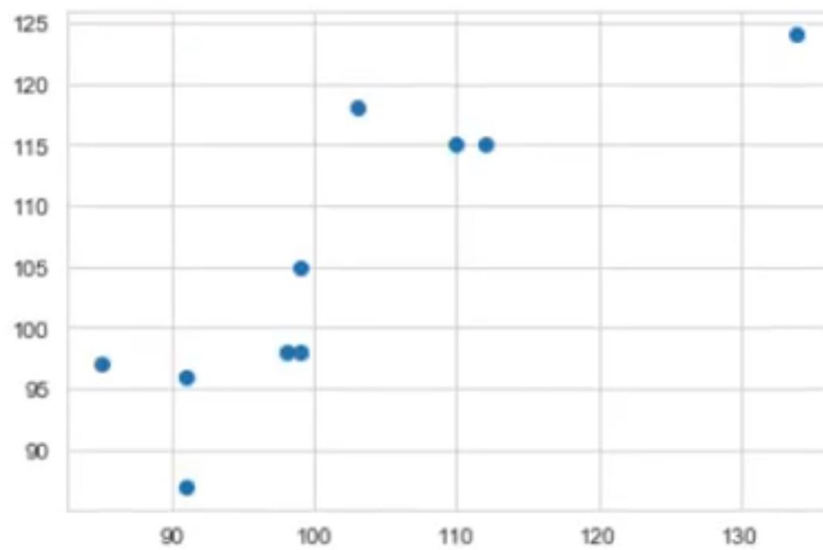
Out[14]: <matplotlib.axes._subplots.AxesSubplot at 0x1c6c28a0f08>



Comparing crop productions in the year 2013 and 2014 using box plot.

OUTPUT 7:

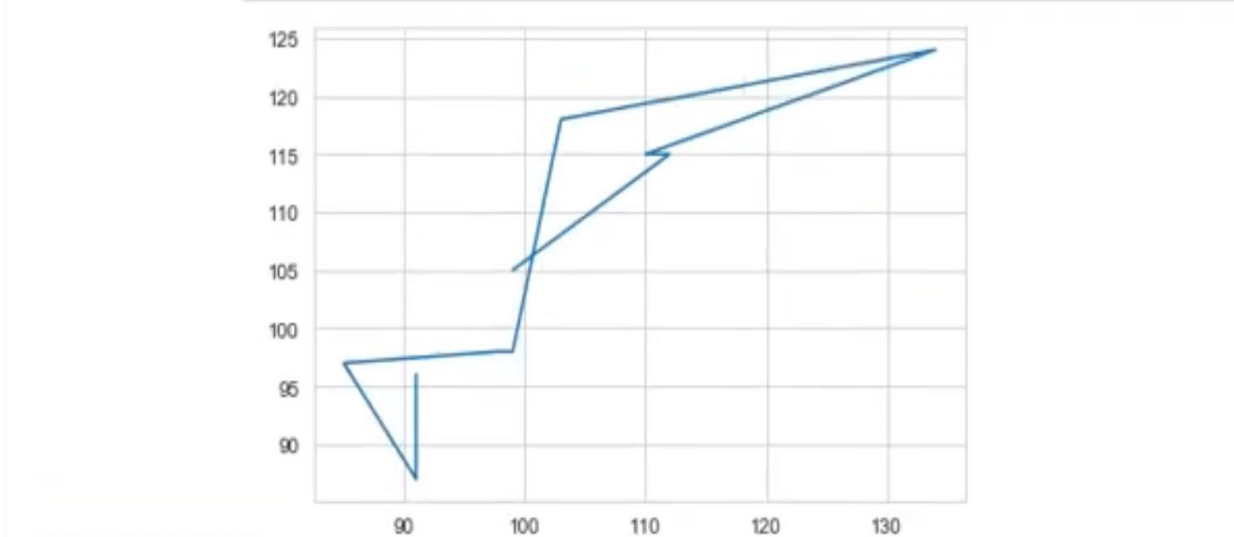

```
In [17]: from matplotlib import pyplot as plt
plt.scatter(data['2013'], data['2014'])
#scatter plot 2013 data vs 2014 data
plt.show()
```



Comparing crop production in the year 2013 and 2014 using scatter plot.

OUTPUT 8:

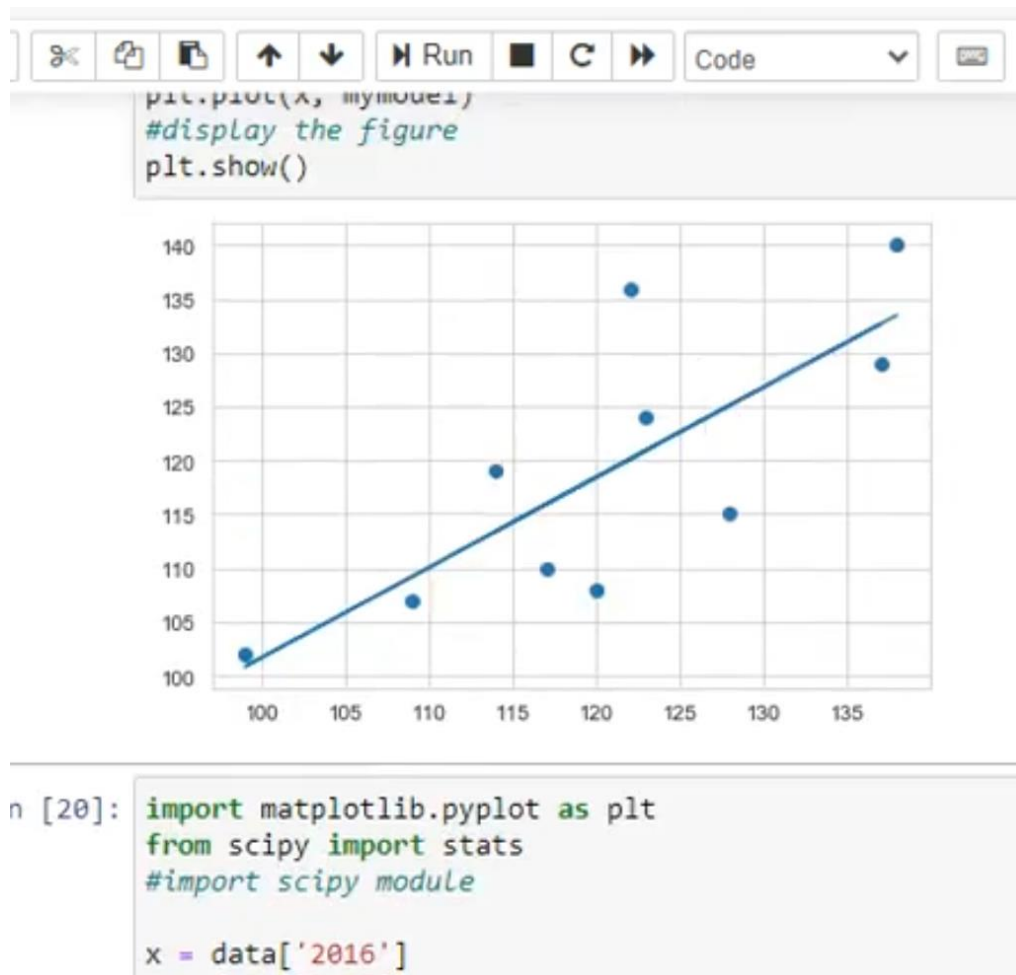
```
In [31]: plt.plot(data['2013'],data['2014'])  
          #line plot 2013 data vs 2014 data  
          plt.show()
```



```
In [19]: import matplotlib.pyplot as plt
          from scipy import stats
          #import scipy module
```

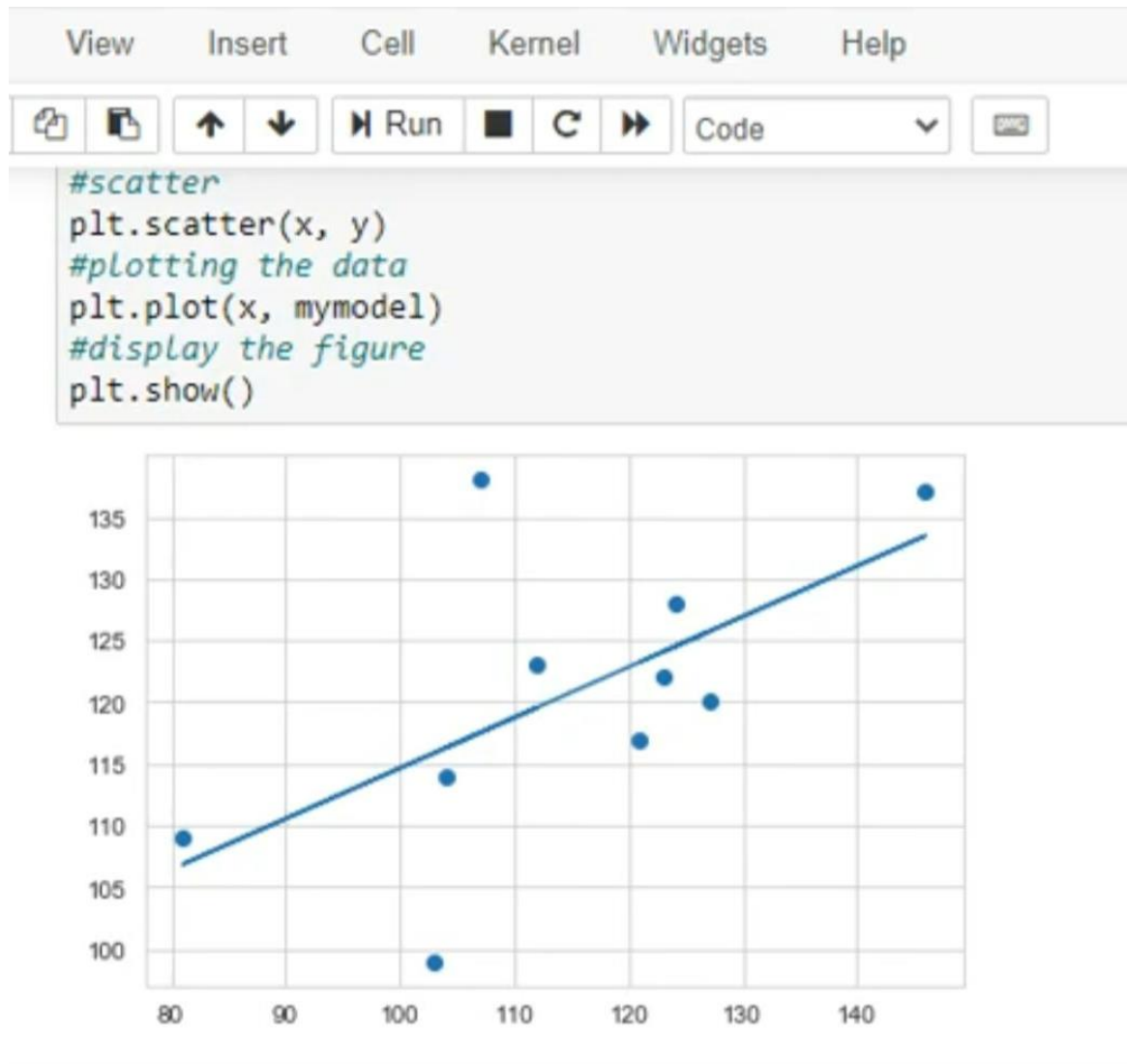
Comparing crop productions in the year 2013 and 2014 using box plot.

OUTPUT 9:



Applying linear regression to visualize and compare predicted crop production data between the year 2017 and 2018.

OUTPUT 10:



Applying linear regression to visualize and compare predicted crop production data between the year 2016 and 2017.

11: CONCLUSION

The proposed system is about the collection of massive dataset collection, the problems in collecting the dataset and also includes collecting the dataset both in the internet as well as in the traditional way.

The massive development in computer technology brings need of most of the data so they produced the system of collecting more number of dataset.

By considering this dataset, a Supply-Demand Prediction forecasting model has been developed in this work that guides the farmers in selecting the appropriate crops to grow.

This in turn suffices the actual demand of the society, minimizing the loss for both farmers as well as consumers at peak times.

This model helps in achieving equilibrium in demand and supply of TOP crops that could effectively solve the current crisis.

12. REFERENCES

1. Department of Agriculture welfare <https://agricoop.nic.in/en/all-india-crop-situation>
2. Top ten India Crops <https://www.mapsofindia.com/top-ten/india-crops/>
3. Indian Crops <https://www.fao.org/india/fao-in-india/india-at-a-glance/en/>
4. Production Of Important Crops in Three Largest Producing State
<https://www.indiabudget.gov.in/economicsurvey/doc/stat/>
5. Production Of Important Crops in Three Largest Producing States
<https://www.safalta.com/blog/major-crops-and-their-states-in-india>
6. Major Cropping Seasons <https://byjus.com/free-ias-prep/major-cropping-seasons-in-india/>
7. Major Crops in India <https://unacademy.com/content/upsc/study-material/indian-geography/major-crops-in-india/>
8. Top 10 Agriculture States <https://www.tractorjunction.com/blog/top-10-agriculture-states-in-india/>
9. Agmarknet (2021) <http://agmarknet.gov.in/PriceTrends/Default.aspx>. Accessed 12 Dec 2021
10. A.T.M.S. Ahamed, N.T. Mahmood, N. Hossain, M.T. Kabir, K. Das, F. Rahman, R.M. Rahman
Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh
2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD 2015 - Proceedings (2015),
11. S.H. Bhojani, N. Bhatt Wheat crop yield prediction using new activation functions in neural network
12. H. Cheng, L. Damerow, Y. Sun, M. Blanke Early yield prediction using image analysis