VISVESVARAYA TECHNOLOGICAL UNIVERSITY



BELAGAVI – 590018, Karnataka

INTERNSHIP REPORT

ON

A Predictive Model for Forecasting Demand and Supply Information of TOP crops

Submitted in partial fulfilment for the award of degree(18CSI85)

BACHELOR OF ENGINEERING IN COMPUTER SCIENCE AND ENGINEERING

Submitted by:

NAME: ARTHIK SHETTY
USN:1MV19CS025



Conducted at **COMPSOFT TECHNOLOGIES**



Sir M Visvesvaraya Institute of Technology Department of Computer Science and Engineering Accredited by NBA, New Delhi

Hunasamaranhalli, Bengaluru

Internship report 2022-2022 1

Sir M Visvesvaraya Institute of Technology Department of Computer Science and Engineering Accredited by NBA, New Delhi

Hunasamaranhalli Bengaluru



This is to certify that the Internship titled "A predictive model for forecasting demand and supply information of TOP crops" carried out by Mr.Arthik shetty, a bonafide student of a Sir M. Visvesvaraya Institute of Technology, in partial fulfillment for the award of Bachelor of Engineering, in computer science and engineering branch under Visvesvaraya Technological University, Belagavi, during the year 2022-2023. It is certified that all corrections/suggestions indicated have been incorporated in the report.

The project report has been approved as it satisfies the academic requirements in respectof Internship prescribed for the course Internship / Professional Practice (18CSI85)

Signature of Guide	Signature of HOD	Signature of Principal
	External Viva:	
Name of the Examiner		Signature with Date
1)		
2)		

Internship report 2022-2022 2

DECLARATION

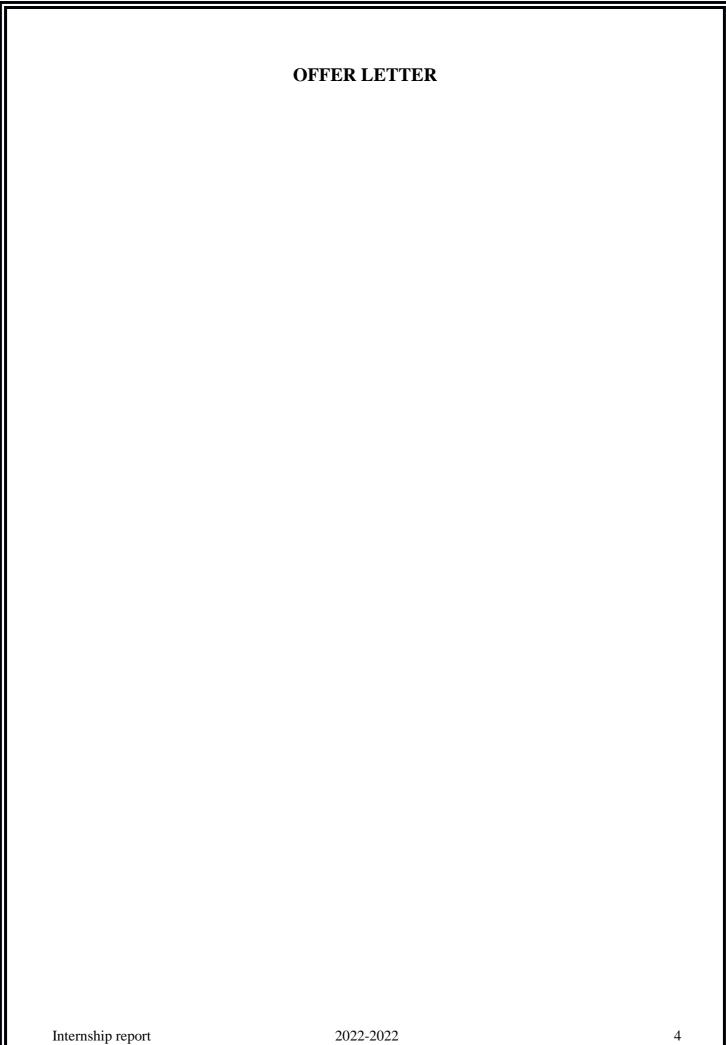
I, **Arthik Shetty** final year student of Computer Science department, Sir M Visvesvaraya Institute of Technology Bengaluru -562157, declare that the Internship has been successfully completed, in **COMPSOFT TECHNOLOGIES**. This report is submitted in partial fulfillment of the requirements for award of Bachelor Degree in Computer Science and Engineering, during the academic year 2022-2023.

Date: 24/09/2022 :

Place: Bengaluru

USN: 1MV19CS025

NAME : Arthik shetty



ACKNOWLEDGEMENT

This Internship is a result of accumulated guidance, direction and support of several

important persons. We take this opportunity to express our gratitude to all who have helped

us to complete the Internship.

We express our sincere thanks to our Principal, **Dr. V.R. Manjunath**, for providing us

adequate facilities to undertake this Internship.

We would like to thank our Head of Dept **Dr. G. C. Bhanu Prakash**, for providing us an

opportunity to carry out Internship and for his valuable guidance and support.

We would like to thank Compsoft Tchnologies for guiding us during the period of

internship.

We would like to thank all the faculty members of our department for the support extended

during the course of Internship.

We would like to thank the non-teaching members of our dept, forhelping us during the

Internship.

Last but not the least, we would like to thank our parents and friends without whose constant

help, the completion of Internship would have not been possible.

NAME: ARTHIK SHETTY

USN: 1MV19CS025

Internship report 2022-2022

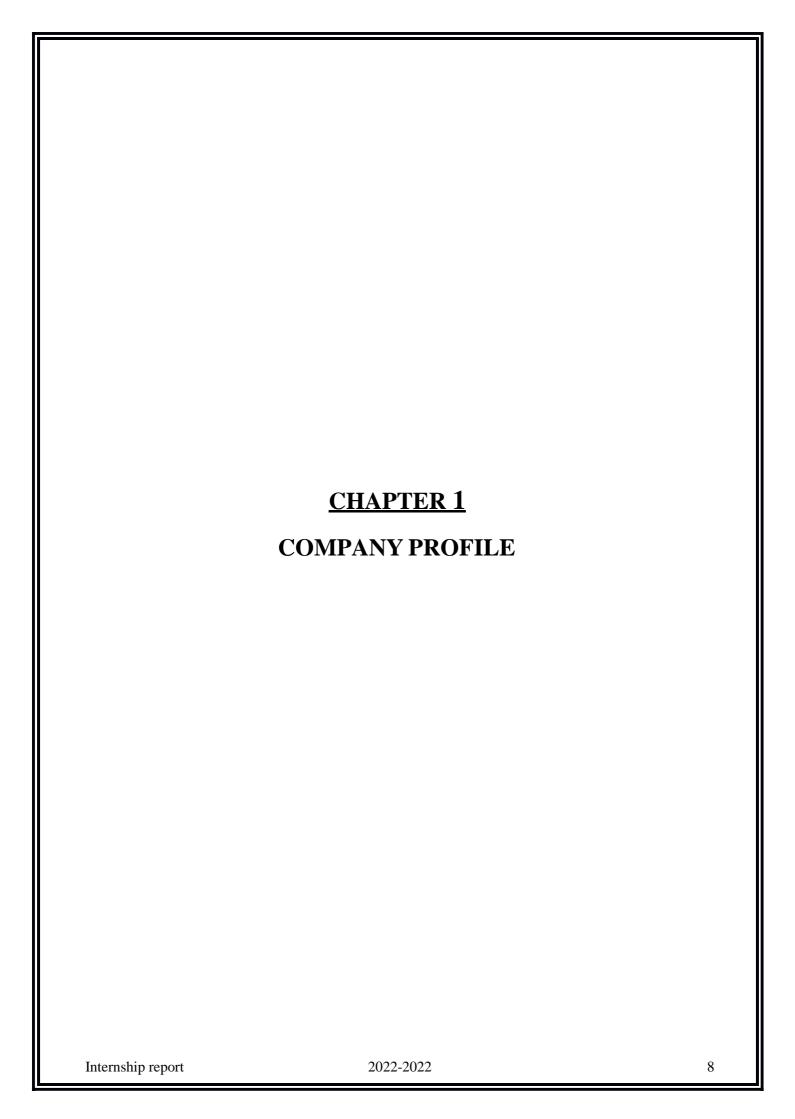
ABSTRACT

India is majorly an agriculture-based economy. 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. The proposed work intends to analyse the various authentic data available for TOP (Tomato, Potato and Onion) crops and design a supply-demand prediction model to forecast the market fluctuations as an advisory measure.

Table of Contents

Sl no	Description	Page no
1	Company Profile	8
2	About the Company	10
3	Introduction	14
4	System Analysis	16
5	Requirement Analysis	18
6	Design Analysis	20
7	Implementation	22
8	Snapshots	24
9	Conclusion	28
10	References	30

Internship report 2022-2022 7



1. COMPANY PROFILE

A Brief History of Compsoft Technologies

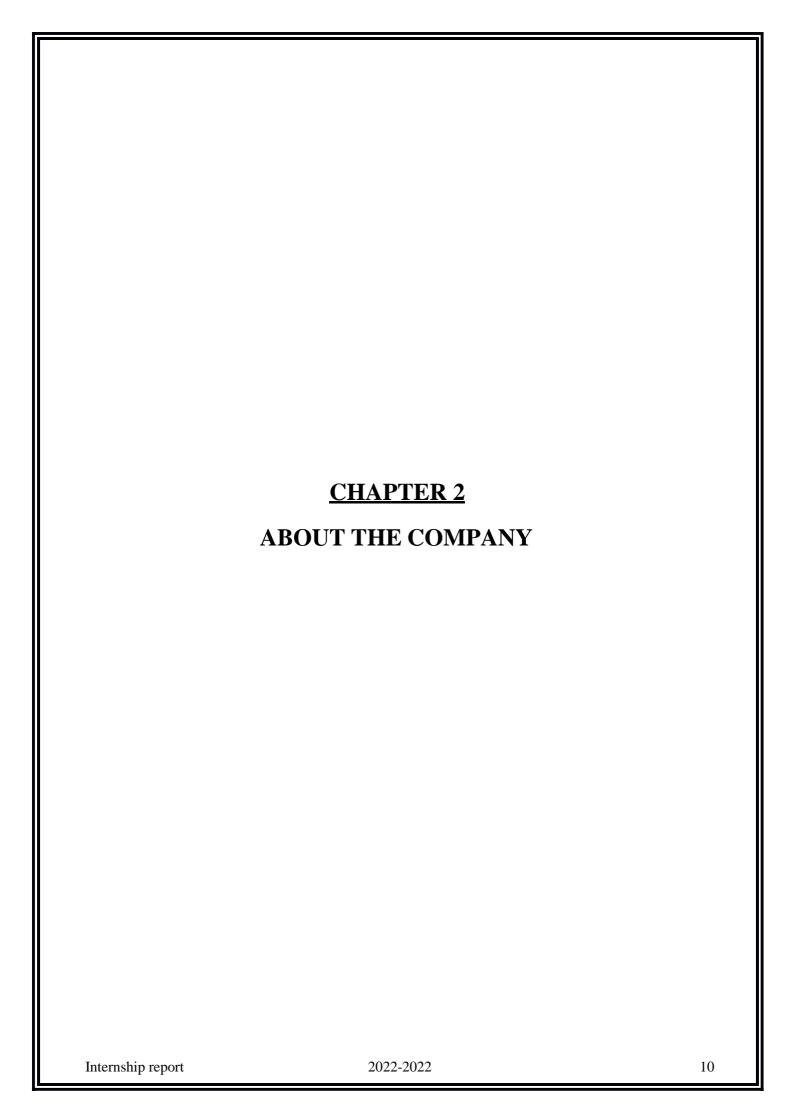
Compsoft Technologies, was incorporated with a goal "To provide high quality and optimal Technological Solutions to business requirements of our clients". Every business is a different and has a unique business model and so are the technological requirements. They understand this and hence the solutions 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 itin one sentence "Technology helps you to Delight your Customers" and that is what we want to achieve.



2. 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 a 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 specifific 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 specifified 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 is 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; and

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 mark up 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.

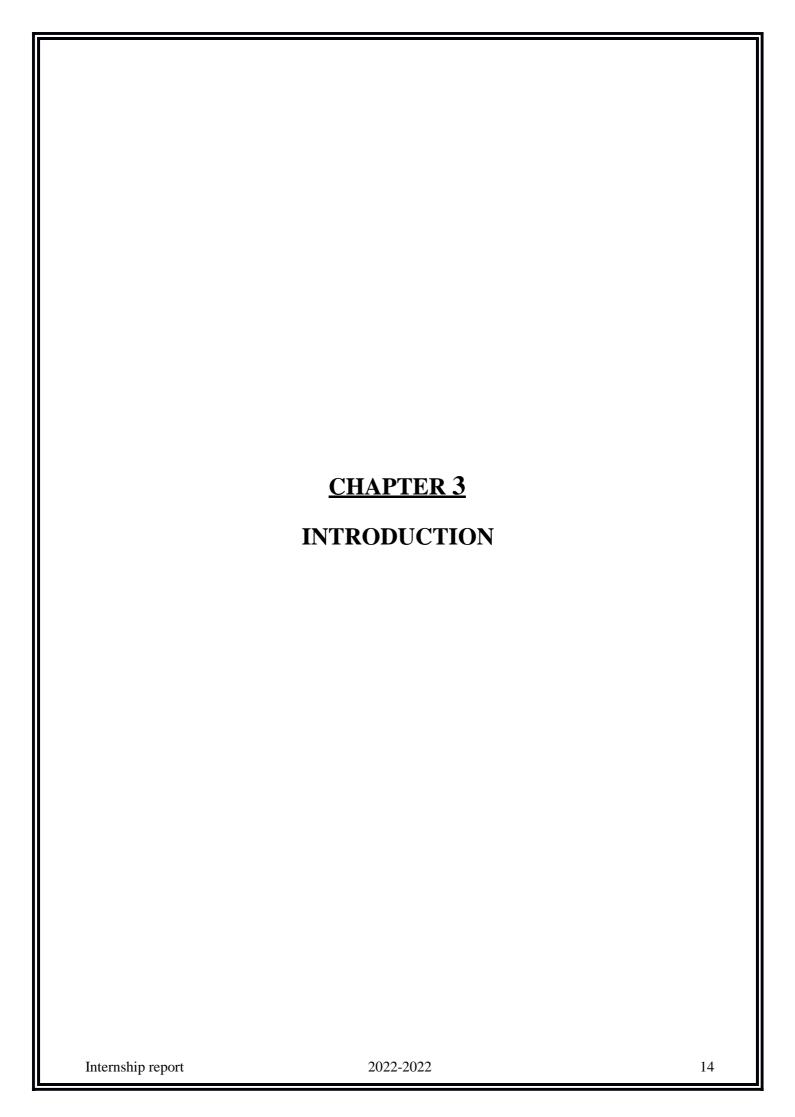
Departments and services offered

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



3. INTRODUCTION

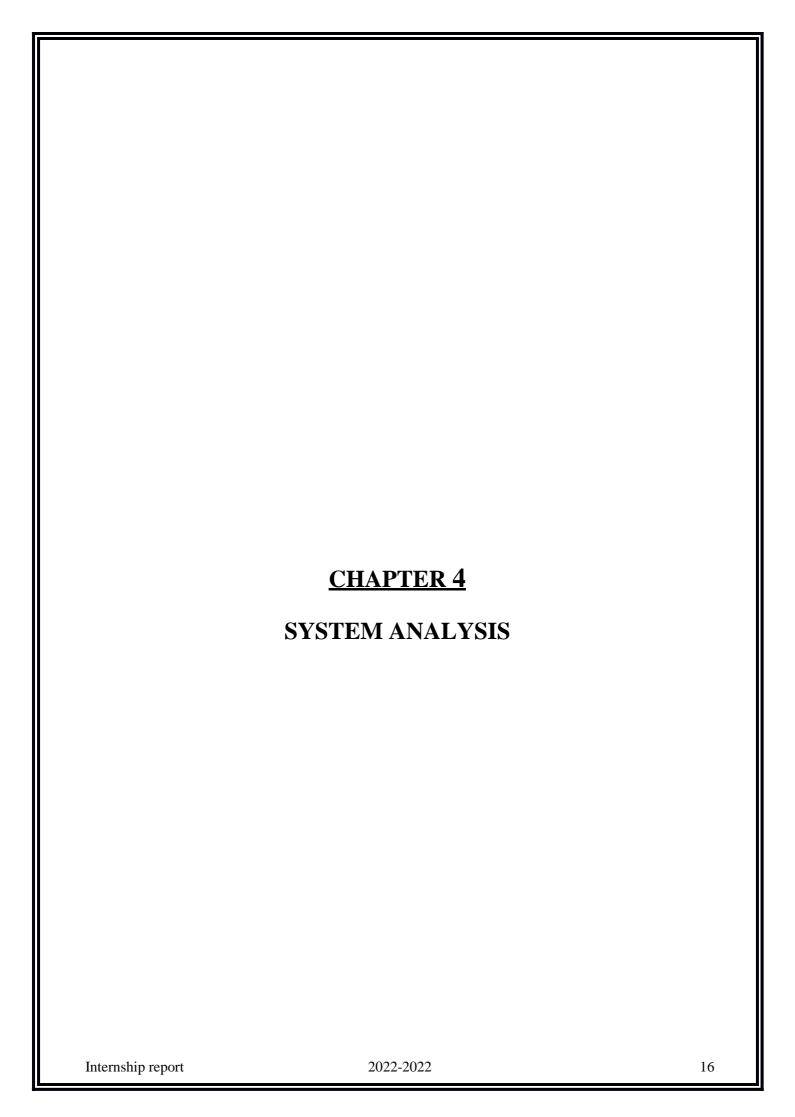
Introduction to ML:

Machine learning is a subfield of artificial intelligence (AI). The goal of machine learning generally is to understand the structure of data and fit that data into models that can be understood and utilized by people.

Although machine learning is a field within computer science, it differs from traditional computational approaches. In traditional computing, algorithms are sets of explicitly programmed instructions used by computers to calculate or problem solve. Machine learning algorithms instead allow for computers to train on data inputs and use statistical analysis in order to output values that fall within a specific range. Because of this, machine learning facilitates computers in building models from sample data in order to automate decision-making processes based on data inputs.

Problem Statement:

India population in 2021 is estimated to be 139 Crore. The increase in population will be more in developing countries like India. The economic growth also has been increased dramatically in the recent years 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 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 Tomato, Potato and onion. 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 commodities set too high then the suppliers (farmers) try 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 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.



4. SYSTEM ANALYSIS

1. Existing System:

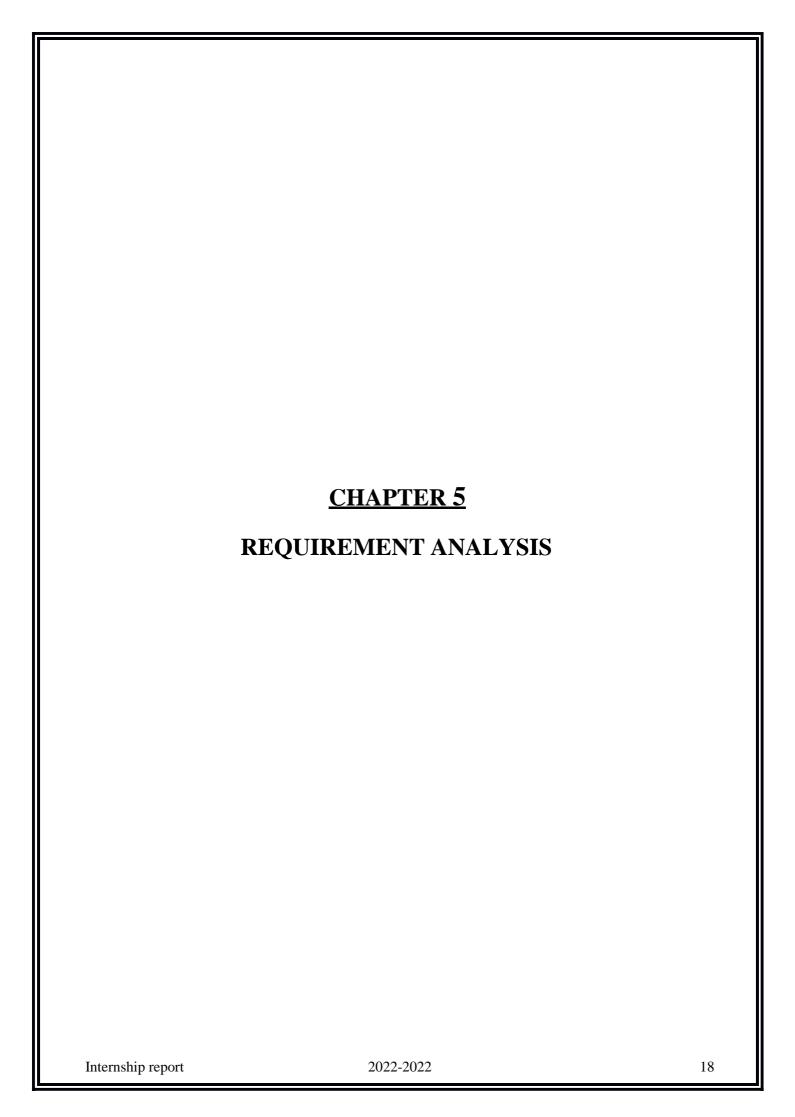
For farmers, the arrangement of a financial motivation is an exceptionally significant factor in choice to plant or reject the manor of yields. Besides, data on demand and supply of TOP crops can provide useful information to farmers, governments, and the common man. It can inform issues about price, demand and production of the crop. The agriculture sector enables 64 % of the rural workforce. According to the latest survey of Agricultural Households conducted by National sample survey office, nearly half of the farmers' income come from the crop production. The critical farming decisions are made by analyzing the price data. The price chart for TOP crops are prepared by gathering, reviewing and analyzing the various data collected from the different authentic sources. Every part of the information is inspected and assessed utilizing diagnostic and coherent thinking. Demand supply analysis can track the price of the crops which will influence the farmers decision of crop production.

2. Proposed System:

This analysis helps the farmers in evaluating future demands of the crops. This report will help farmers determine the variety and the time of planting the crops. Main objective of this work includes helping the farmers by providing historical crop yield data with cost forecasting for risk management. Also, data collected would help the administration in making crop protection arrangements and strategies for inventory network activity. This will likewise enable the government to have an equilibrium costing over the TOP crops, so they can sell the item in the market for reasonable expense.

3. Objective of the System:

The work was carried out in four stages- data collection, data pre-processing, prediction and data visualization. General parameters for predicting the crop price are climate change, Government policy and demand. The acquired data was not present in the required format; thus, it was structured in the required format. Prediction is done by utilizing past year data of crop price. The output provides one year data i.e., a yearly price prediction. Consider two input parameters for yearly price prediction. The price of the crop which is output under a certain period, it is the price of input in the previous period. Machine Learning algorithms are used to predict and forecast demand-supply and prices. Prediction is accomplished using regression analysis.



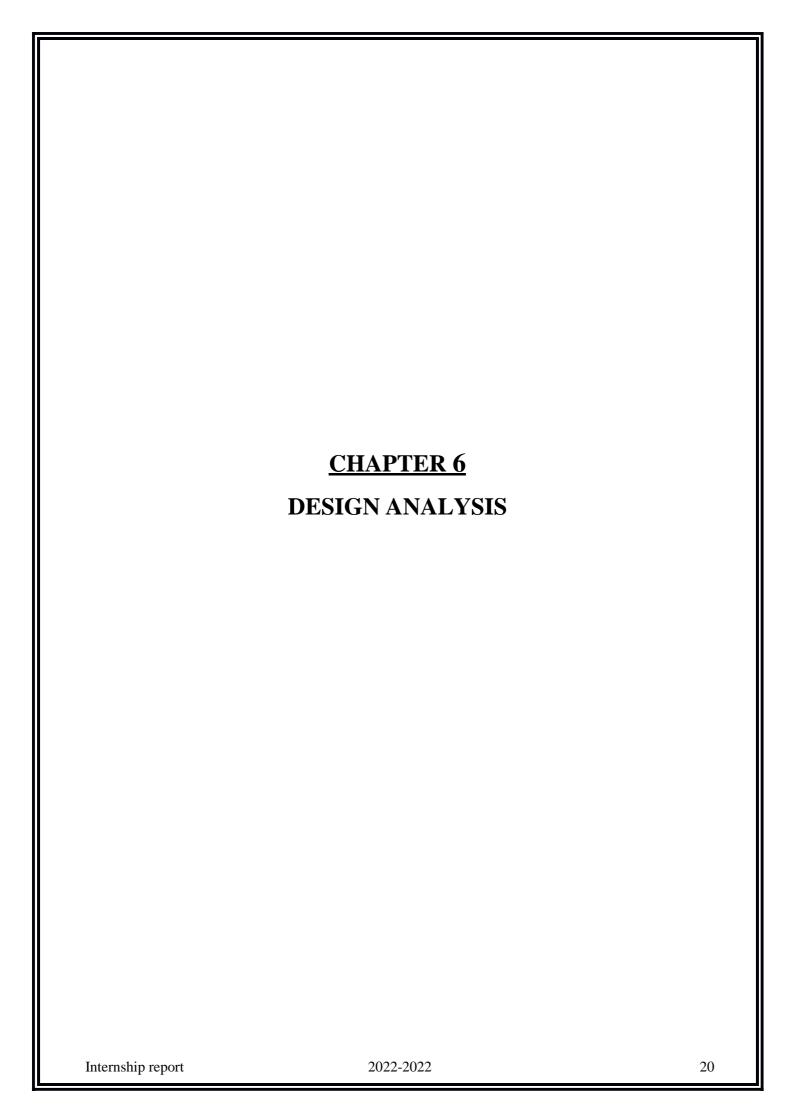
5. REQUIREMENT ANALYSIS

Hardware Requirement Specification:

- Ubuntu 16.04 or later (64-bit)
- macOS 10.12.6 (Sierra) or later (64-bit) (no GPU support)
- Windows 7 or later (64-bit)
- Raspbian 9.0 or later.
- GPU support requires a CUDA®-enabled card (Ubuntu and Windows)

Software Requirement Specification

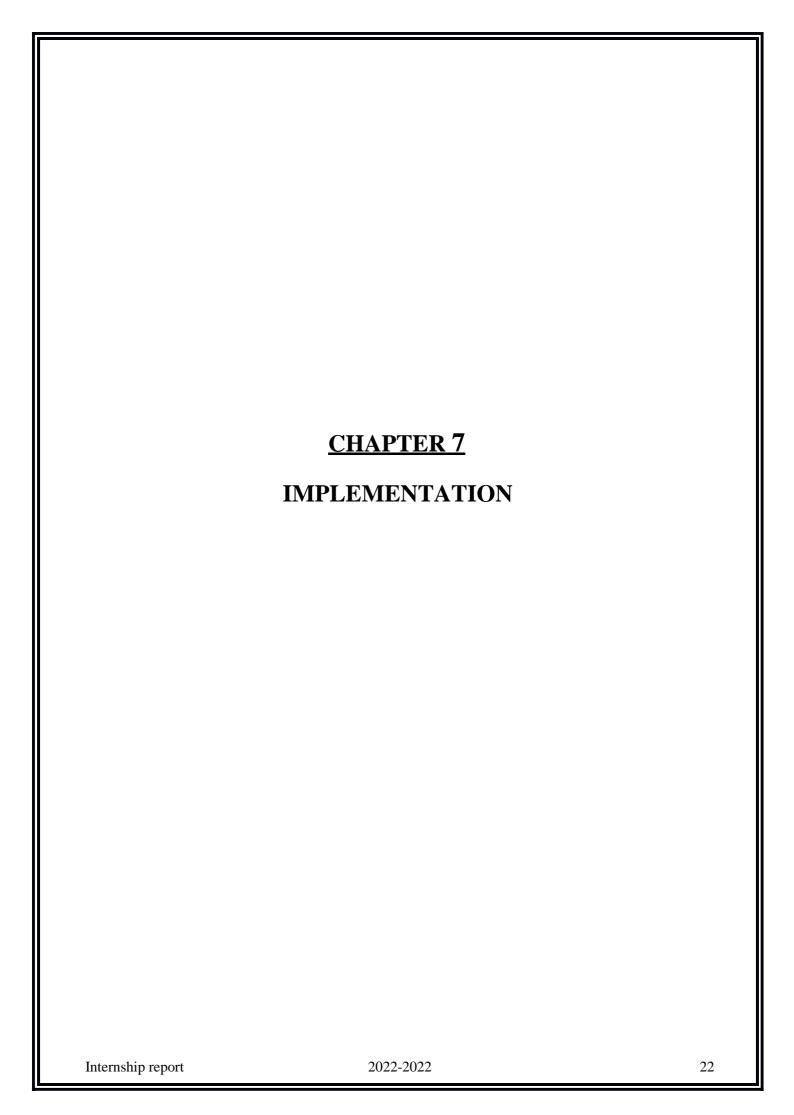
- Python 3.5–3.7.
- pip 19.0 or later (requires manylinux2010 support)
- Tensorflow **2.10. 0**



6. DESIGN & ANALYSIS

The work was carried out in four stages- data collection, data pre-processing, prediction and data visualization. General parameters for predicting the crop price are climate change, Government policy and demand. The acquired data was not present in the required format; thus, it was structured in the required format. Prediction is done by utilizing past year data of crop price. The output provides one year data i.e., a yearly price prediction. Consider two input parameters for yearly price prediction. The price of the crop which is output under a certain period, it is the price of input in the previous period. Machine Learning algorithms are used to predict and forecast demand-supply and prices. Prediction is accomplished using regression analysis. It is a type of supervised learning in the domain of Machine Learning that results in a predicted relationship between labels and data points.

- 1. Data acquisition and ingestion: 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, Indiastat.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.
- 2. Data pre-processing: 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.
- **3. Designing the predictive model:** 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.
- **4. Model Validation:** 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.



7. 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 it constraints on implementation, design of methods to achieve the change over and an evaluation of change over methods 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.

	CHAPTER 8	
	SNAPSHOTS	
Internship report	2022-2022	24

8. SNAPSHOTS

Datasets:

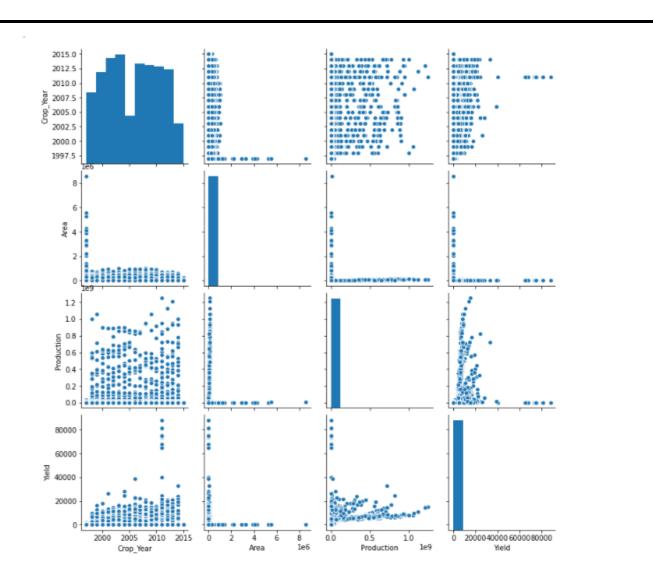
	State_Name	District_Name	Crop_Year	Season	Crop	Area	Production
0	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Arecanut	1254.0	2000.0
1	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Other Kharif pulses	2.0	1.0
2	Andaman and Nicobar Islands	NICOBARS	2000	Kharif	Rice	102.0	321.0
3	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Banana	176.0	641.0
4	Andaman and Nicobar Islands	NICOBARS	2000	Whole Year	Cashewnut	720.0	165.0

246086	West Bengal	PURULIA	2014	Summer	Rice	306.0	801.0
246087	West Bengal	PURULIA	2014	Summer	Sesamum	627.0	463.0
246088	West Bengal	PURULIA	2014	Whole Year	Sugarcane	324.0	16250.0
246089	West Bengal	PURULIA	2014	Winter	Rice	279151.0	597899.0
246090	West Bengal	PURULIA	2014	Winter	Sesamum	175.0	88.0

246091 rows × 7 columns

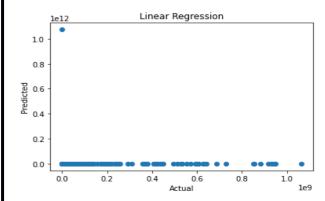
	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice
2195	107	34	32	26.774637	66.413269	6.780064	177.774507	coffee
2196	99	15	27	27.417112	56.636362	6.086922	127.924610	coffee
2197	118	33	30	24.131797	67.225123	6.362608	173.322839	coffee
2198	117	32	34	26.272418	52.127394	6.758793	127.175293	coffee
2199	104	18	30	23.603016	60.396475	6.779833	140.937041	coffee

2200 rows × 8 columns



```
plt.scatter(y_test,lr_predict)
plt.xlabel('Actual')
plt.ylabel('Predicted')
plt.title('Linear Regression')
```

Text(0.5, 1.0, 'Linear Regression')



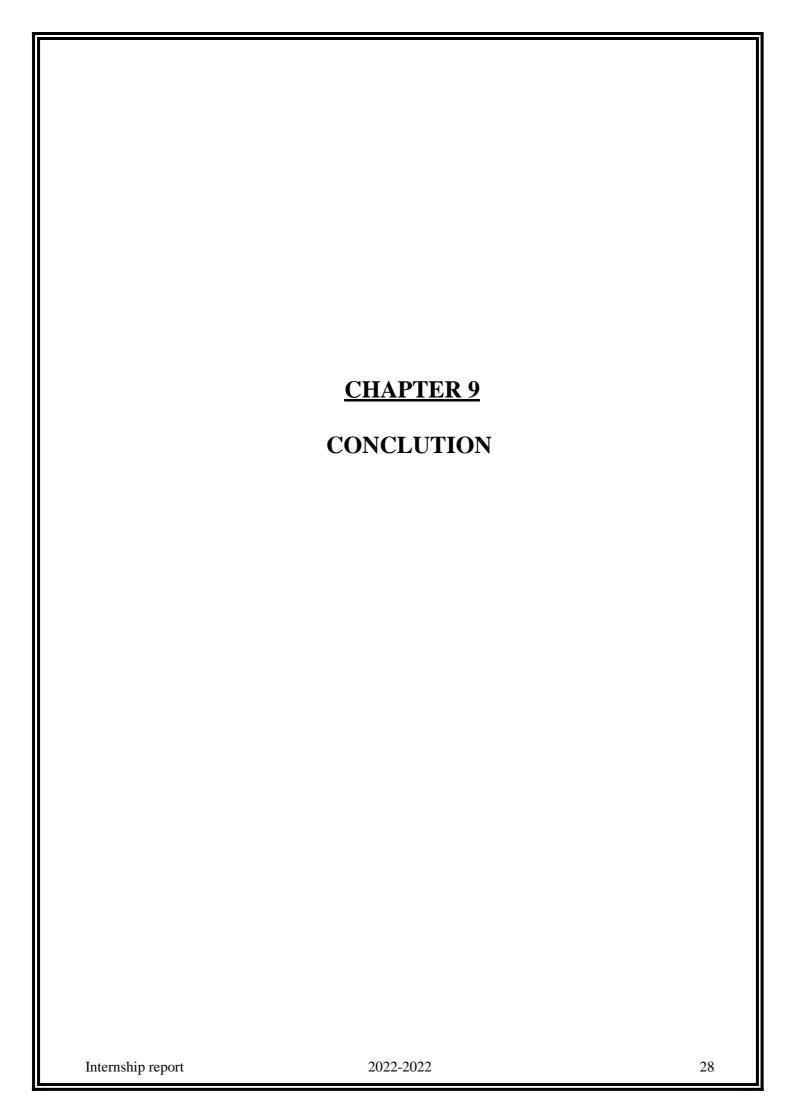
Clearly, the dataset is not good for linear regression.

Assumptions of Linear Regression

- 1. Linearity.
- 2. Homoscedasticity
- 3. Multivariate normality
- 4. Lack of multicollinearity

This tells us how well the unknown samples will be predicted by our model.

Random Forest Algorithm



9. CONCLUTION

The package was designed in such a way that future modifications can be done easily. The following conclusions can be deduced from the development of the project:

- ❖ Automation of the entire system improves the efficiency
- ❖ It provides a friendly graphical user interface which proves to be better when compared to the existing system.
- **!** It gives appropriate access to the authorized users depending on their permissions.
- ❖ It effectively overcomes the delay in communications.
- Updating of information becomes so easier
- System security, data security and reliability are the striking features.
- ❖ The System has adequate scope for modification in future if it is necessary.

10. REFERENCE

- 1. Xindong Wu et. al., "Top 10 algorithms in data analysis". Knowledge and Information Systems, vol. 14, pp. 1-37- Jan. 2018. [1][3]
- **2.** Guo-Qiang, Luo Chang-shou, Wei Qing-Feng, "Prediction and Research on Vegetable Price based Genetic Algorithm and Neural network model" Asia Agricultural Research, Vol. 3, No. 5, pp. 148-150, 2011.[4]
- **3.** Divya Chauhan, Jawahar Thakur, "Data analytic Techniques for Weather Prediction: A Review", International Journal on Recent and Innovation Trends in Computing and Communication, Vol.2 No. 8, pp. 2184 2189- 2016.[4]
- **4.** Mucherino, P. Papajorgji, P.M. Pardalos, "A Survey of Data Mining Techniques applied to Agriculture", Operational Research, Vol. 9, No. 2, pp. 121-140- 2019.[4][5]
- **5.** Pravar Jain "Artificial Intelligence in Agriculture: Using Modern Day AI to Solve Traditional Farming Problems" in Data Science Blogathon- November 2020.[2]
- 6. Darshan Halliyavar, Shwetha M P, Sharanabasavaraj, Puneeth S and Shivaprasad Ashok Chikop "Production Analysis and Prediction of TOP using ARIMA Model (PAPTOP)" in 2020 International Journal of Engineering Research & Technology (IJERT) PP 1-4 (IEEE)-2020.[3][4][5]
- 7. Agmarknet (2021) http://agmarknet.gov.in/PriceTrends/Default.aspx. Accessed 12 Decc2021.
- **8.** Hopcoms(2021) http://www.hopcoms.kar.nic.in/(S(qbaji2453cfnhs55umapkyzk))/ default.aspx.
- **9.** Department of Agriculture Farmers Welfare(2021)https://agricoop.nic.in/en/all-india-cropsituation.
- 10. NITI Aayog (2021) https://www.niti.gov.in/verticals/data-management-and-analysis.

