

**INDUSTRY INTERNSHIP**  
with  
**UNICONVERGE TECHNOLOGIES PVT. LTD.**



In partial fulfilment of the requirements for the award of the degree  
OF  
Bachelor's in Technology  
IN  
Computer Science  
BY  
**Prabal Kalhans**

Under the guidance of

**Mr. Kaushlendra Singh Sisodia**  
[Founder]

## Internship Offer Letter



### UpSkill Campus

C-56/11, Ground floor, Near IT Stellar Park, Sector -62, Noida, U.P. – 201309  
Email: [support@upskillcampus.com](mailto:support@upskillcampus.com), Ph. No. 9354068856

### Internship Offer Letter

Date: **14-May-2023**

Name: **Prabal Kalhans**

Email: **Kalhans.prabal@gmail.com**

Mob. **919910000000**

Dear Candidate,

We are pleased to inform you that UpSkill Campus alongwith it's Industry Partner **UniConverge Technologies Pvt. Ltd.** has agreed to provide you internship

as **Python**

Your appointment of services is based on the following terms and conditions.

1. The internship shall start from **15-May-2023**
2. Internship: Internship period will be for 6 weeks starting from date of joining.
3. Working Hours:  
The regular working hours of the company is 10:00 am to 6.00 pm (Monday-Friday).

On behalf of both the Company, we wish you every success in your position and trust that our relationship will be long and mutually rewarding.

Sincerely  
For UpSkill Campus and UniConverge Technologies Pvt. Ltd.



**Prabha Singh**  
HR Manager

## ABSTRACT

This report presents a comprehensive analysis of agricultural crop production in India from 2001 to 2014, utilizing data science techniques. The primary objective of this project is to predict suitable crops for specific agricultural regions and provide alternative crop recommendations. Through the implementation of various machine learning models, the accuracy of these predictions was evaluated.

The dataset used for this analysis contains extensive information on agricultural crop production, including factors such as crop yield, cost of production, rainfall, temperature, soil quality, and geographical location. These features were analyzed to identify patterns and trends in crop production across different regions and time periods. Initially, exploratory data analysis (EDA) techniques were employed to gain insights into the dataset. Visualizations and statistical analyses were utilized to understand the relationships between various features and crop production. The findings from EDA guided the subsequent steps of the project, including feature engineering, preprocessing, and model selection.

To predict suitable crops for specific agricultural regions, three machine learning models were implemented: Decision Tree Classifier, Random Forest Classifier, and Logistic Regression, but only one is mentioned because of confidential reasons. These models were trained on a subset of the dataset, with features such as rainfall, temperature, soil quality, and geographical location. The accuracy of these models was evaluated using a testing dataset, and the best performing model was selected for further analysis. The selected model was then utilized to predict suitable crops for a particular area, based on the given features. Additionally, the model was designed to recommend alternative crops, providing valuable insights to farmers and agricultural stakeholders. The accuracy of these predictions was assessed, enabling an understanding of the model's reliability and effectiveness.

The results obtained from this project demonstrate the potential of data science techniques in analyzing agricultural datasets and predicting suitable crops for specific regions in India. The findings can be used to aid decision-making processes in the agricultural sector, optimizing crop selection and promoting sustainable agriculture practices.

*Keywords:* Data science, agricultural crop production, machine learning, prediction, suitability analysis, alternative crop recommendations, accuracy evaluation.

## **CONTENTS**

Abstract.....	iii
Contents.....	v
<b>1 INTRODUCTION.....</b>	<b>5</b>
1.1 The Problem.....	5
1.2 Factors associated.....	5
1.3 About the Dataset.....	5
1.4 More about the Dataset.....	6
1.5 Additional Information.....	6
1.6 Background of the Project.....	7
1.7 Machine Learning in Agriculture.....	7
1.8 Dynamic Nature of weather.....	7
<b>2 METHODOLOGY.....</b>	<b>9</b>
2.1 Data Collection.....	9
2.2 Identifying outliers.....	9
2.3 Data Analysis.....	9
<b>3 RESULT.....</b>	<b>18</b>
<b>4 FUTURE SCOPE.....</b>	<b>19</b>
<b>5 REFERENCES.....</b>	<b>20</b>

## **CHAPTER-1: INTRODUCTION**

### **1.1 The Problem:**

Agriculture has long been a cornerstone of human civilization, providing sustenance and nourishment for communities around the world. In a country like India, agriculture holds paramount importance, serving as the backbone of the nation's economy and a source of livelihood for millions of people. With its diverse climatic conditions, vast agricultural landscapes, and rich agricultural heritage, India has played a significant role in the global agricultural sector. In this report, we delve into an extensive dataset encompassing various aspects of crop production in India from 2001 onwards. The dataset includes information on different crop types, years of production, quantity of production, cultivation area, and the associated costs of cultivation and production across different states of India. India's agricultural sector stands as a testament to the nation's rich history and deep-rooted connection to the land. Over the years, agriculture has been the lifeblood of India's economy, providing employment, sustenance, and cultural significance. As the world's second-most populous country, the role of agriculture in India is not limited to domestic concerns alone. The country's agricultural prowess has far-reaching implications on global food security, trade, and environmental sustainability.

### **1.2 Factors associated with the problem statement:**

With a diverse range of agro-climatic zones, India boasts an extensive variety of crops that thrive in different regions. From the fertile plains of Punjab to the coastal fields of Kerala, each state contributes its unique agricultural offerings to the nation's overall production. The dataset under analysis encompasses this diverse tapestry of crops, documenting their types, production quantities, cultivation areas, and associated costs. Such comprehensive information provides a valuable resource for understanding the intricacies of India's agricultural landscape. India's prominence in the agricultural sector stems from its significant contributions to global food production, as well as its immense agricultural potential. The country ranks second in the world in terms of arable land, making it an agricultural powerhouse. Moreover, agriculture in India employs a substantial portion of the population, especially in rural areas, where agriculture is the primary occupation. The agricultural sector not only provides employment but also plays a vital role in ensuring food security and poverty alleviation.

### **1.3 About the dataset:**

The dataset under study presents a comprehensive picture of the agricultural landscape in India, capturing crucial factors that influence crop production and sustainability. It encompasses data on various crops cultivated in different regions of India, shedding light on the geographic diversity of agricultural practices within the country. The dataset further provides insights into the annual production quantities of crops, measured in tonnes and quintals, offering a quantitative understanding of crop output over the years. The dataset's inclusion of production quantities measured in tonnes and quintals allows for a comprehensive analysis of the nation's crop output over time. Examining production trends and patterns enables us to identify factors influencing fluctuations in yield, such as climate variability, technological advancements, and market demands. Furthermore, analyzing the cultivation area, measured in hectares, reveals the spatial distribution of crops across the country. This information sheds light on regional specialization, resource allocation, and the impacts of land-use patterns on agricultural productivity.

#### 1.4 More about the dataset:

Within the dataset, the incorporation of cost-related data provides essential insights into the economic dimensions of crop cultivation. Assessing the cost of cultivation and production across different states helps in understanding the financial challenges faced by farmers and the potential barriers to agricultural sustainability. By analyzing cost dynamics, policymakers and stakeholders can develop strategies to optimize input usage, improve efficiency, and enhance the livelihoods of farmers. In addition to production quantities, the dataset also encompasses information on the cultivation area in terms of hectares. This aspect is of great significance as it enables the assessment of crop productivity and resource allocation in different regions. Analyzing the cultivation area can provide valuable insights into the efficiency and effectiveness of agricultural practices, guiding policymakers and farmers towards optimal land use and crop selection strategies.

The dataset also includes data on the cost of cultivation and production in different states of India. Understanding the cost dynamics in the agricultural sector is crucial for evaluating profitability, economic viability, and sustainability of farming practices. By analyzing the cost data, we can identify factors that impact the profitability of crop cultivation, such as input costs, labor expenses, and market conditions. This knowledge can help in formulating strategies to enhance agricultural productivity and address challenges faced by farmers. Taking this analysis a step further, the project aims to harness the power of data science and machine learning to predict suitable crops for specific agricultural regions. By training and evaluating various machine learning models, we seek to develop accurate prediction algorithms that can guide farmers in selecting crops most suitable for their geographic location, climate conditions, and available resources. Such predictions have the potential to increase agricultural productivity, reduce risks, and contribute to the sustainable development of India's agricultural sector.

#### 1.5 Additional information:

Additionally, the project aims to provide alternative crop recommendations. By leveraging the dataset's comprehensive information, we can identify crops that offer viable alternatives to existing practices, promoting crop diversification and resilience. These recommendations are crucial in the face of climate change, evolving market dynamics, and the need for sustainable agricultural practices that safeguard both productivity and the environment. In this report, our objective is not only to explore the dataset and analyze the trends and patterns in crop production but also to leverage the power of data science and machine learning to predict the suitable crops for specific agricultural regions. By employing various machine learning models, we aim to develop accurate predictions that can assist farmers and policymakers in making informed decisions regarding crop selection. Furthermore, we seek to provide alternative crop recommendations based on the dataset, enabling diversification and resilience in agricultural practices.

By undertaking this comprehensive analysis and prediction endeavor, we aspire to contribute to the advancement of agricultural practices in India. Through informed decision-making, data-driven strategies, and the empowerment of farmers and policymakers, we strive to unlock the true potential of India's agricultural sector. Ultimately, this project aims to foster a future where sustainable agricultural practices, food security, and economic prosperity go hand in hand, not only within India's borders but also as a beacon of agricultural excellence on the global stage. Through this comprehensive analysis and prediction of crop suitability, we aim to contribute to the advancement of agricultural practices in India. By harnessing the insights gleaned from this dataset, we hope to promote sustainable agriculture, optimize resource allocation, and bolster food security. Ultimately, this report endeavors to shed light on the crucial role of agriculture in

India, not only for the nation itself but also in the global context, as we navigate the challenges of feeding a growing population while preserving our planet's resources.

#### 1.6 Background of the project:

This paper focuses on utilizing machine learning techniques to predict crop types and calculate their yield accurately. Various machine learning methodologies were employed to determine the accuracy of the predictions. Specifically, the Random Forest classifier was utilized to predict the suitable crops for a selected district. The implementation of a system for crop prediction based on historical data collection was accomplished. This system aims to assist farmers in making informed decisions regarding crop selection for their fields. The proposed technique provides valuable insights that enable efficient and productive harvesting practices. By accurately predicting the most suitable crops for different districts, farmers in Kerala can benefit from increased yields. Ultimately, this work contributes to the growth of the Indian economy by maximizing crop production rates.

#### 1.7 Machine Learning in Agriculture:

This study examines a collection of publications that explore yield prediction using machine learning techniques. The selected papers employ various features based on the research scope and data availability. Each paper investigates yield prediction but differs in the specific features considered. Additionally, variations in scale, geographical location, and crop type contribute to the diversity of approaches. The selection of features is contingent upon dataset availability and research objectives. Notably, studies indicate that models with more features do not consistently outperform those with fewer features in yield prediction. To determine the best-performing model, it is crucial to test models with different feature sets. Multiple algorithms have been employed across different studies, and while no definitive conclusion can be drawn regarding the best model, certain machine learning models emerge as more commonly used. Notably, the random forest, neural networks, linear regression, and gradient boosting tree models are frequently employed. Several studies employ multiple machine learning models to assess their predictive performance. Given the prevalence of neural networks, this study also investigates the use of deep learning algorithms for crop yield prediction. Upon examining 30 relevant papers, the most utilized deep learning algorithms were found to be CNN, LSTM, and DNN. However, other algorithms have also been applied to this problem. This article serves as a foundation for further research on the development of crop yield prediction, shedding light on the current state of the field and paving the way for future advancements.

#### 1.8 Dynamic nature of the weather vs yield:

Crop yields are heavily influenced by weather conditions, and understanding this relationship is crucial for assessing the potential impacts of climate change on the agricultural sector. Numerous studies have been conducted to model this relationship and project the effects of climate change on crop yields. In this paper, we present an innovative approach to yield modeling that combines the strengths of deep neural networks with semiparametric techniques. This approach allows us to capture complex nonlinear relationships in high-dimensional datasets while incorporating known parametric structures and accounting for unobserved cross-sectional heterogeneity. These findings emphasize the value of our novel approach in accurately capturing the complex dynamics between weather patterns and crop yields. By accounting for both parametric and nonparametric aspects, our model provides a more nuanced understanding of climate change impacts on

corn yield. This has important implications for policymakers, farmers, and stakeholders involved in climate change adaptation and mitigation strategies in the agricultural sector.

## CHAPTER-2: METHODOLOGY

### 2.1 Data Collection:

As far as data collection is concerned, gathered a comprehensive dataset containing information on crop types, years of production, quantity of production, cultivation area, cost of cultivation, and other relevant variables for agricultural crops in India from 2001-2014. Ensure the dataset covers diverse geographical regions and crop varieties. Cleaning the dataset by removing any duplicates, missing values, or outliers. Normalize or standardize numerical variables if necessary was then done. To identify any pattern or understand any inner insight of the dataset, it is important that we gather a deeper understanding by exploring it more which was then subsequently done.

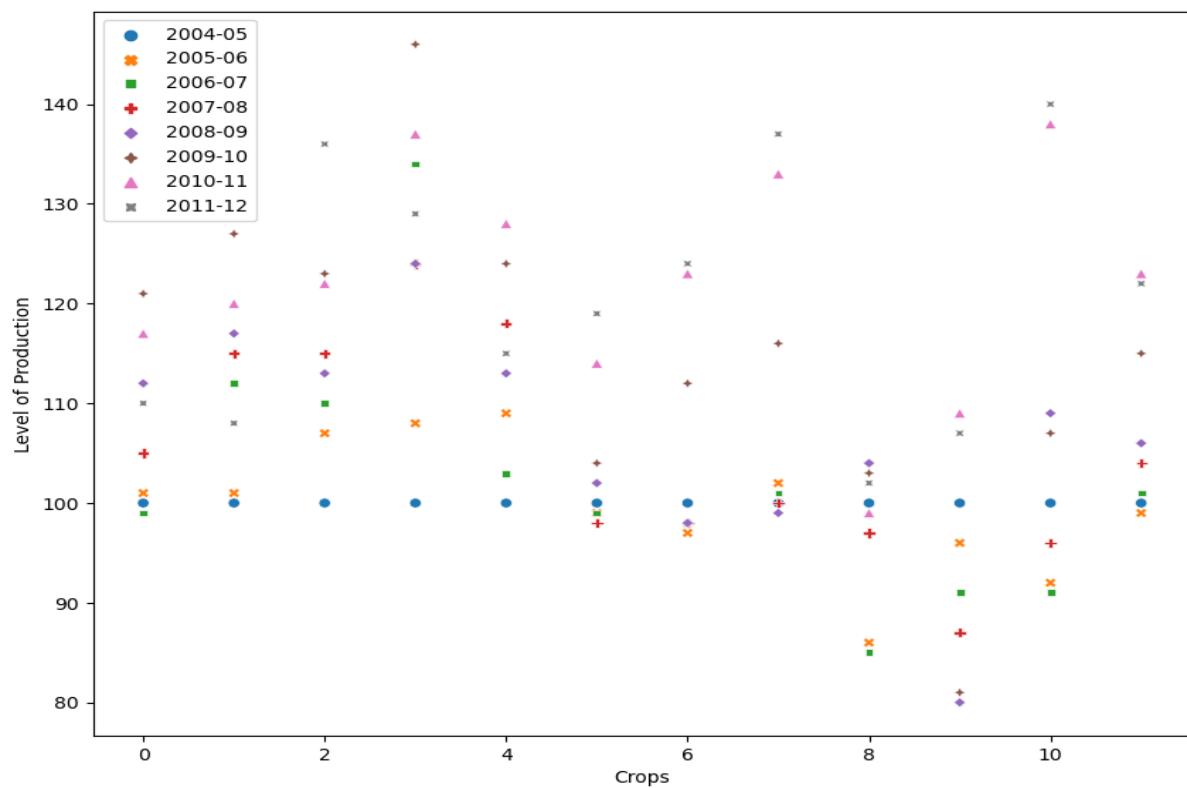


Figure 2.1: Help determine any outlier

### 2.2 Data Analysis and EDA:

Exploratory Data Analysis Techniques need to be performed for examining the relationship between variables, identify trends, correlations, and distribution within the dataset. Visualization tools in the python language can really become handy in such cases where a deeper understanding of the dataset is required.

To know, how the dataset is distributed within itself, a boxplot can be handy which was also used for the better understanding.

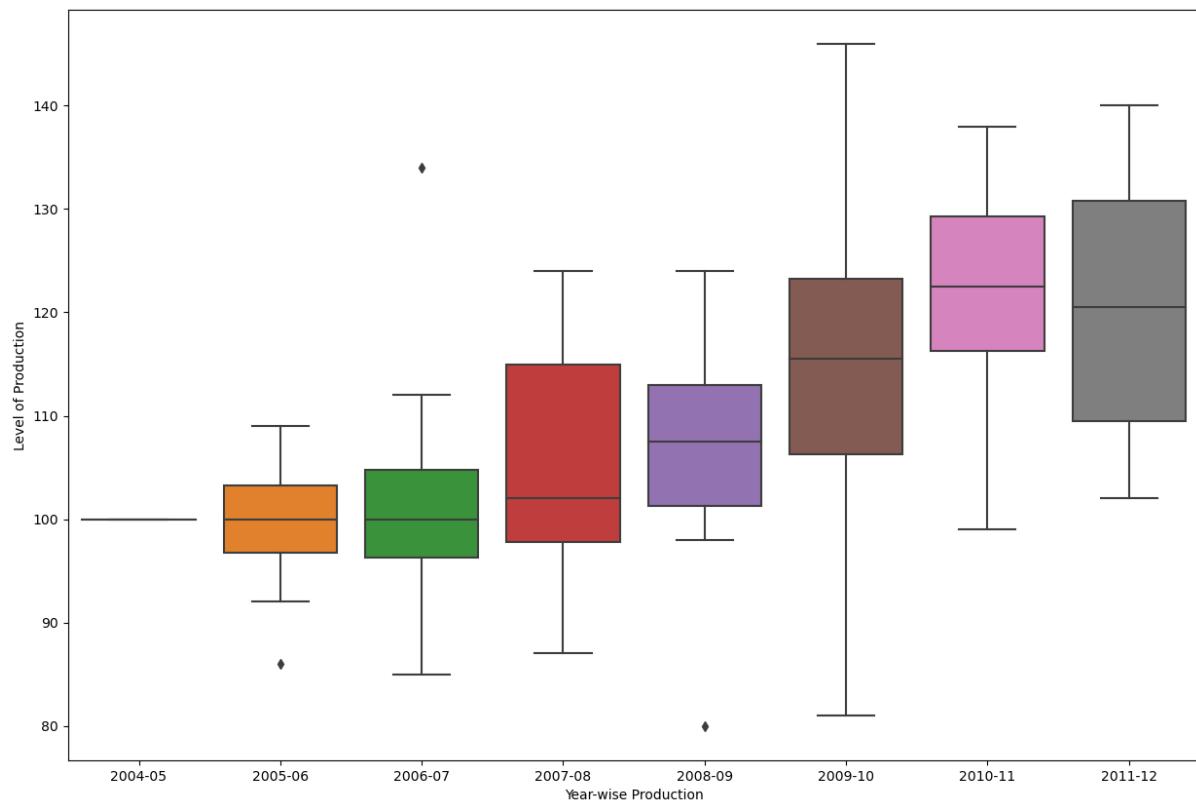


Figure 1.2: Boxplot, to understand distribution

A simple plot can be used to show what crop was produced in which quantity and in what quantity.

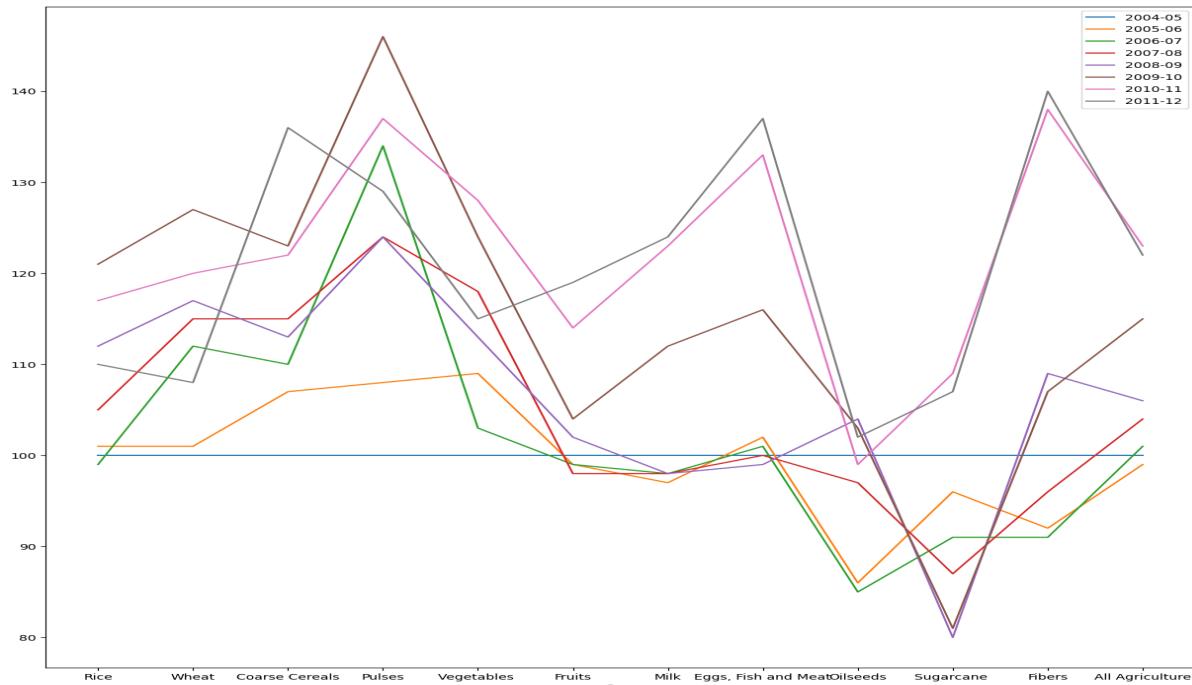


Figure 2.2: Plot to understand the produce over years

It is important to understand that India has a good number of states and that too with different climatic conditions all together. Hence, a statewise understanding of what crops are cultivated and produced holds a significant importance to us.

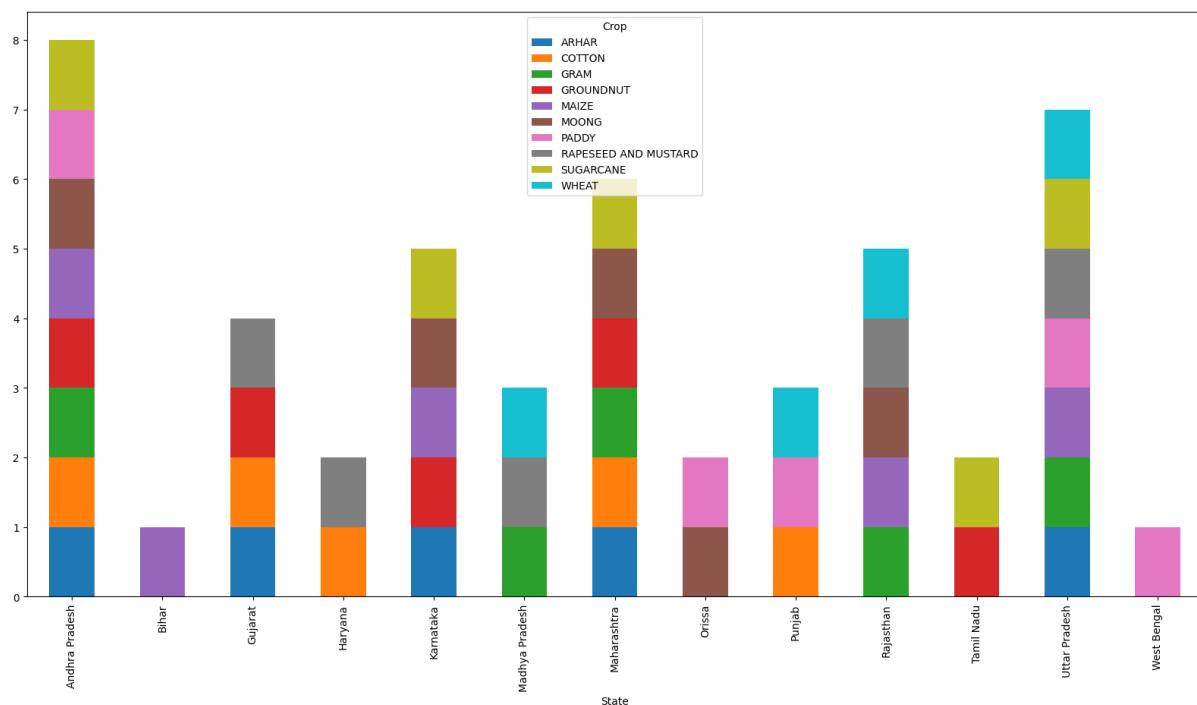


Figure 2.3: Bar graph, to understand the state-wise produce of different crops

### 2.3 Comparative Analysis:

It's very important to realise that a country with more than 140+ crore people, the area vs yield of crop has to be checked at regular intervals. Therefore, to understand this well, though there were four graphs for one financial year, a combined graph was made to showcase the Area vs Yield.

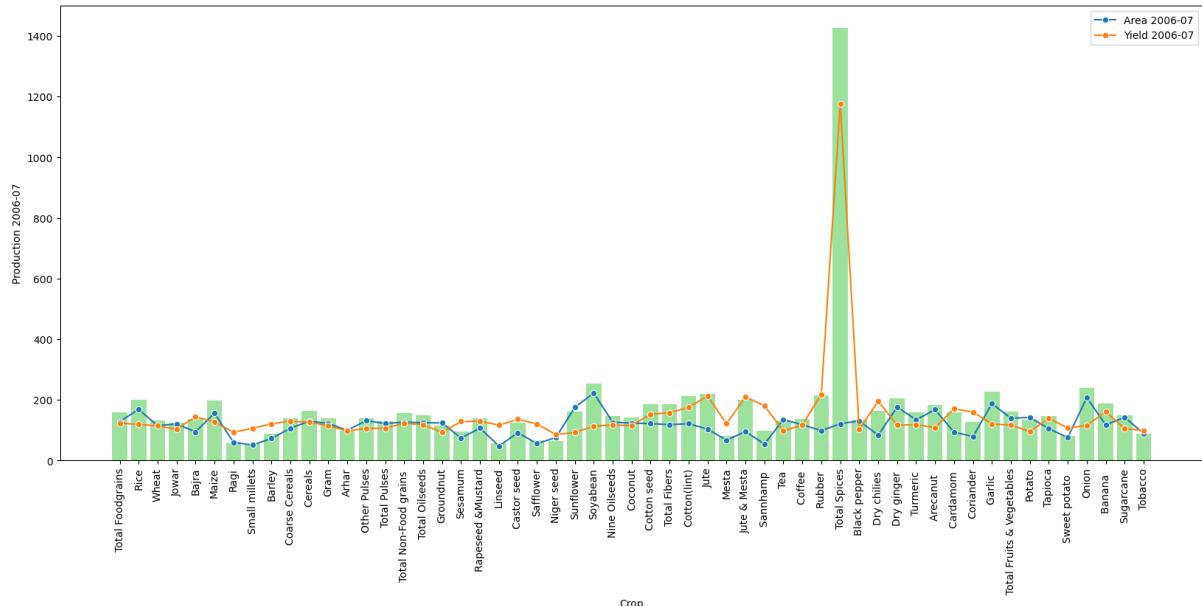


Figure 2.4: Area vs Yield in 2006-2007

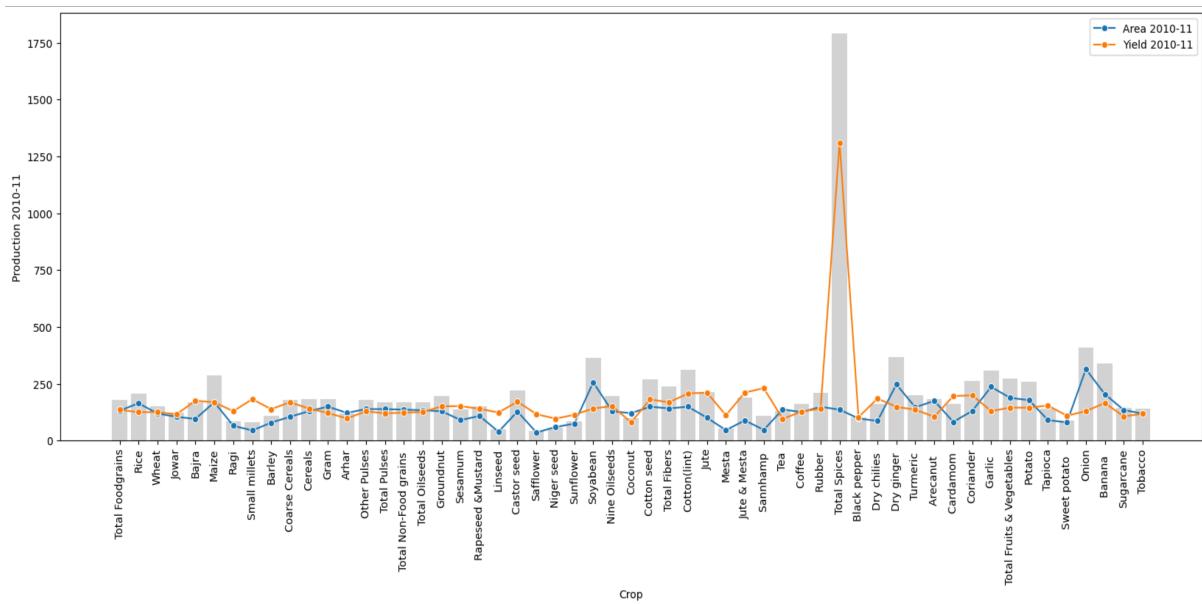


Figure 2.5: Area vs Yield in 2010-2011

With the rapid increase in population over the years, a comparative study needs to be done, about how the area, yield and production have increased or decreased over the years, for every individual crop if possible. The below graphs show the pattern of increase or decrease.

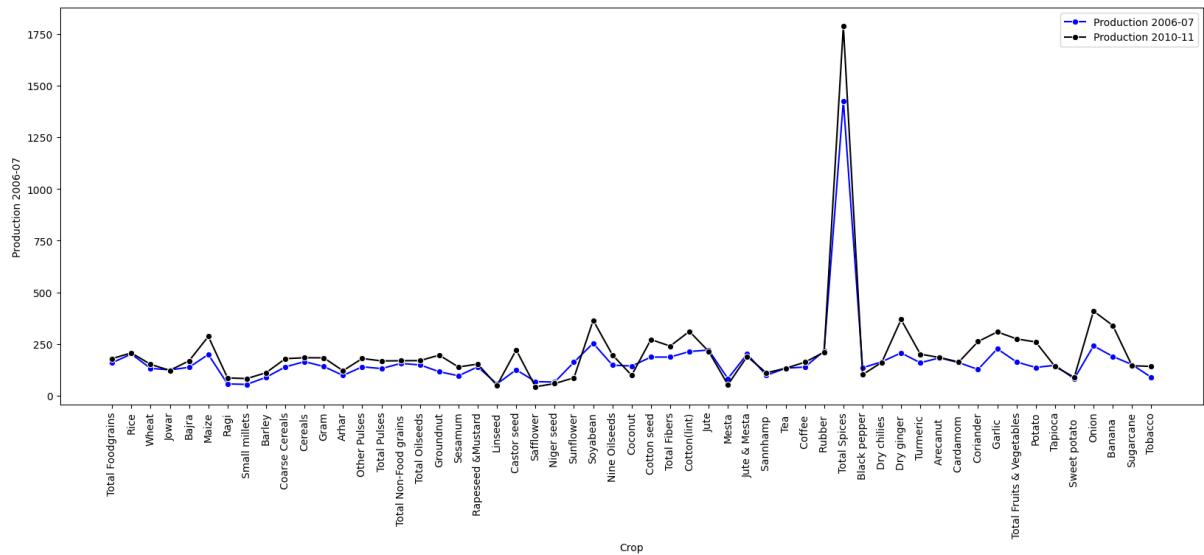


Figure 2.6: Production in 2006-07 vs 2010-11

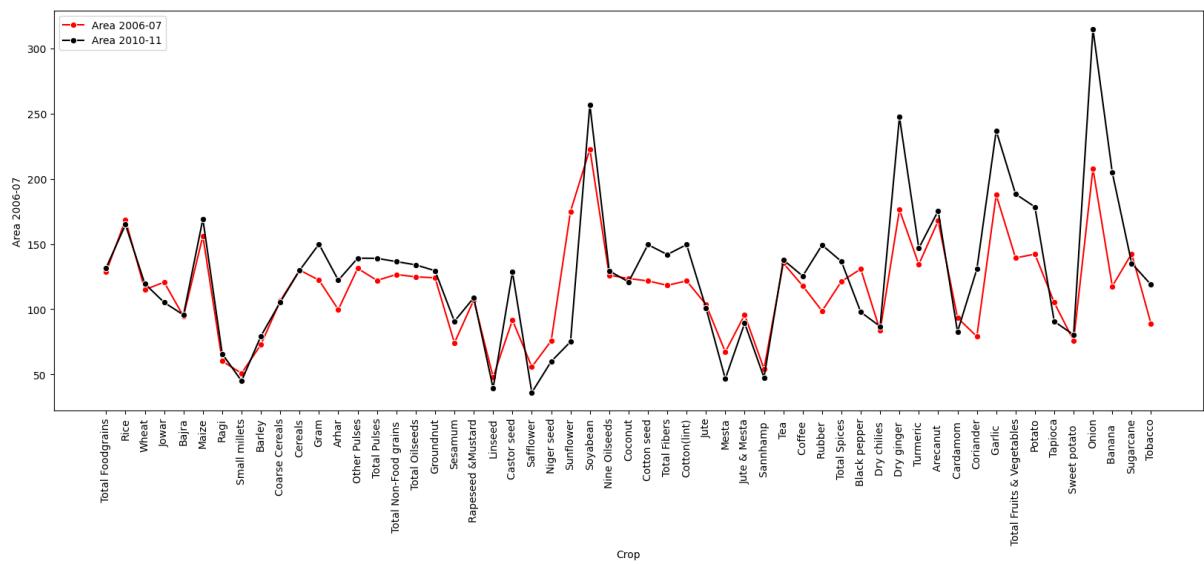


Figure 2.7: Land utilized in year 2006-07 vs 2010-11

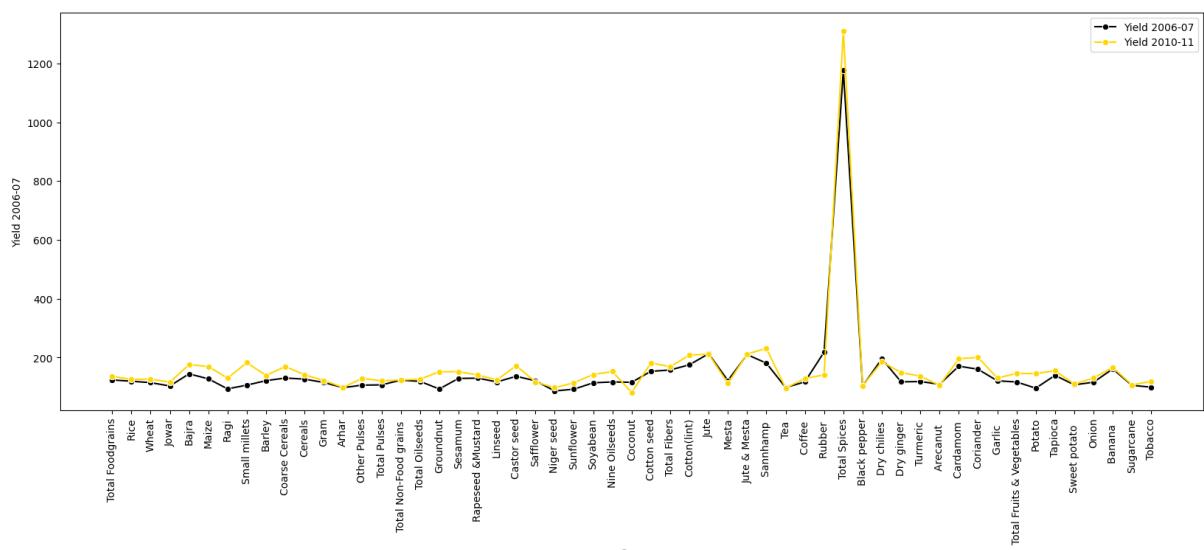


Figure 2.8: Yield in the year 2006-06 vs 2010-11

The graphs above can also be used while understanding why the control in population is required, as there are two things that are somewhat in contrast as of today, urbanisation and growth in agriculture, because to feed a continuously growing population we need larger areas, but again because of increasing population, we need places to live.

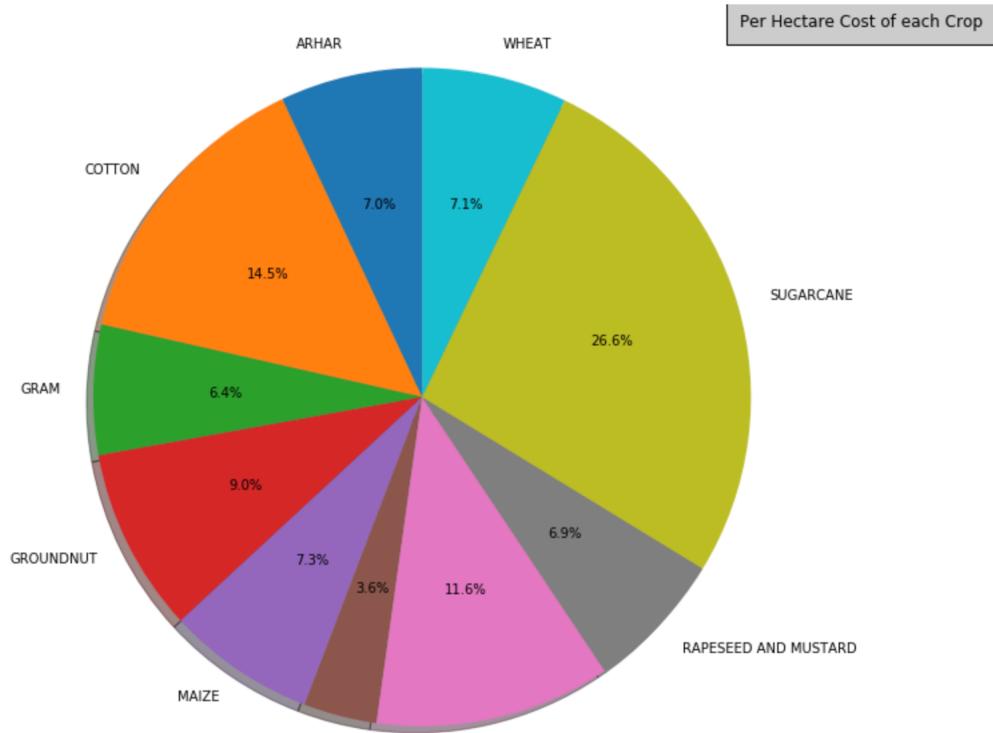


Figure 2.9: Per hectare cost of major crops

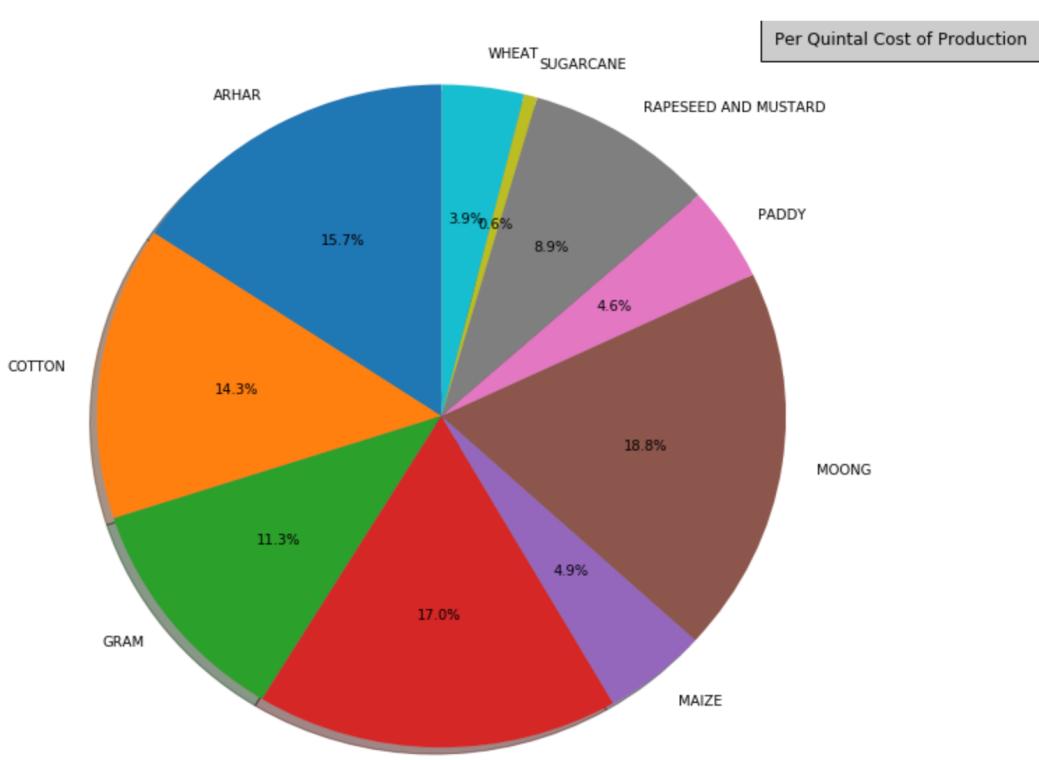
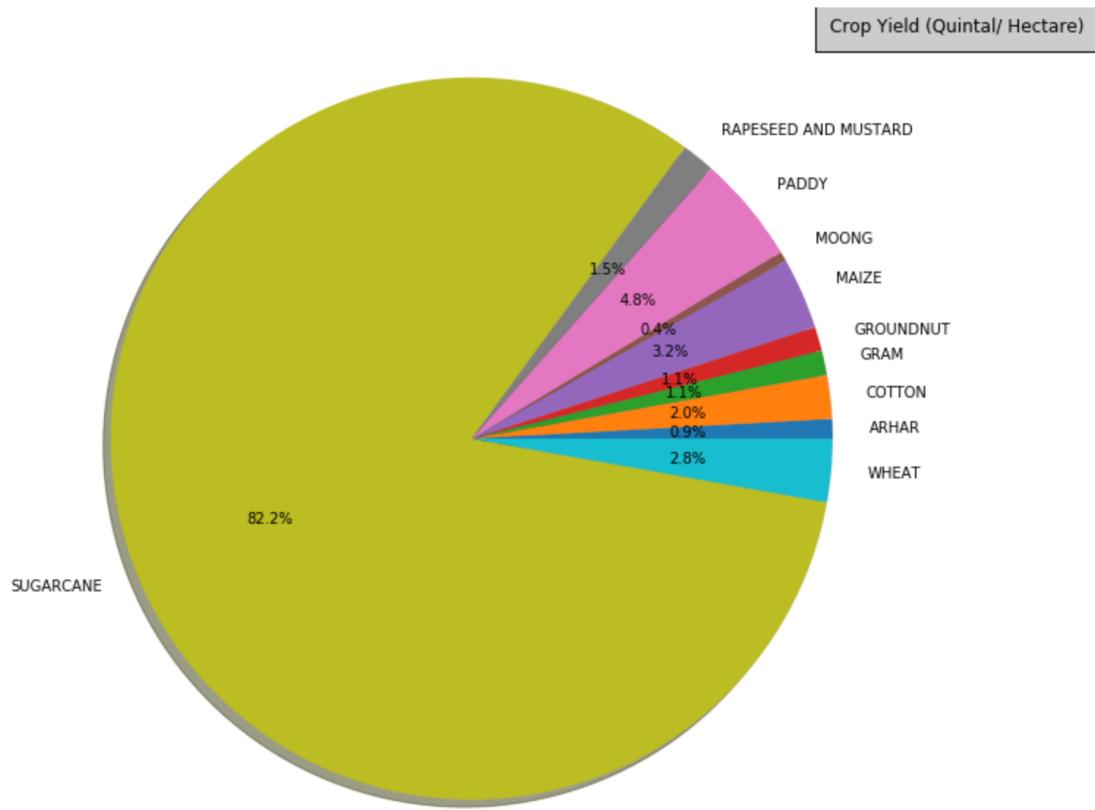


Figure 210: Per quintal cost of production



*Figure 2.11: Crop yield*

From the pie charts that have been displayed above, certain useful information can be drawn which include:

**\*Crop Cultivation-**

- Sugarcane has a highest Yield in India (82.2%)
- Moong has a lowest Yield in India (0.437%)

**\*Cultivation cost for crop-**

- Highest for sugarcane
- Lowest for moong
- Average for 'paddy, maze, cotton, wheat' etc.

**\*Production cost of crop-**

- Sugarcane has very minimal production cost
- Moong has the highest production cost and hence, lowest yield in the country.

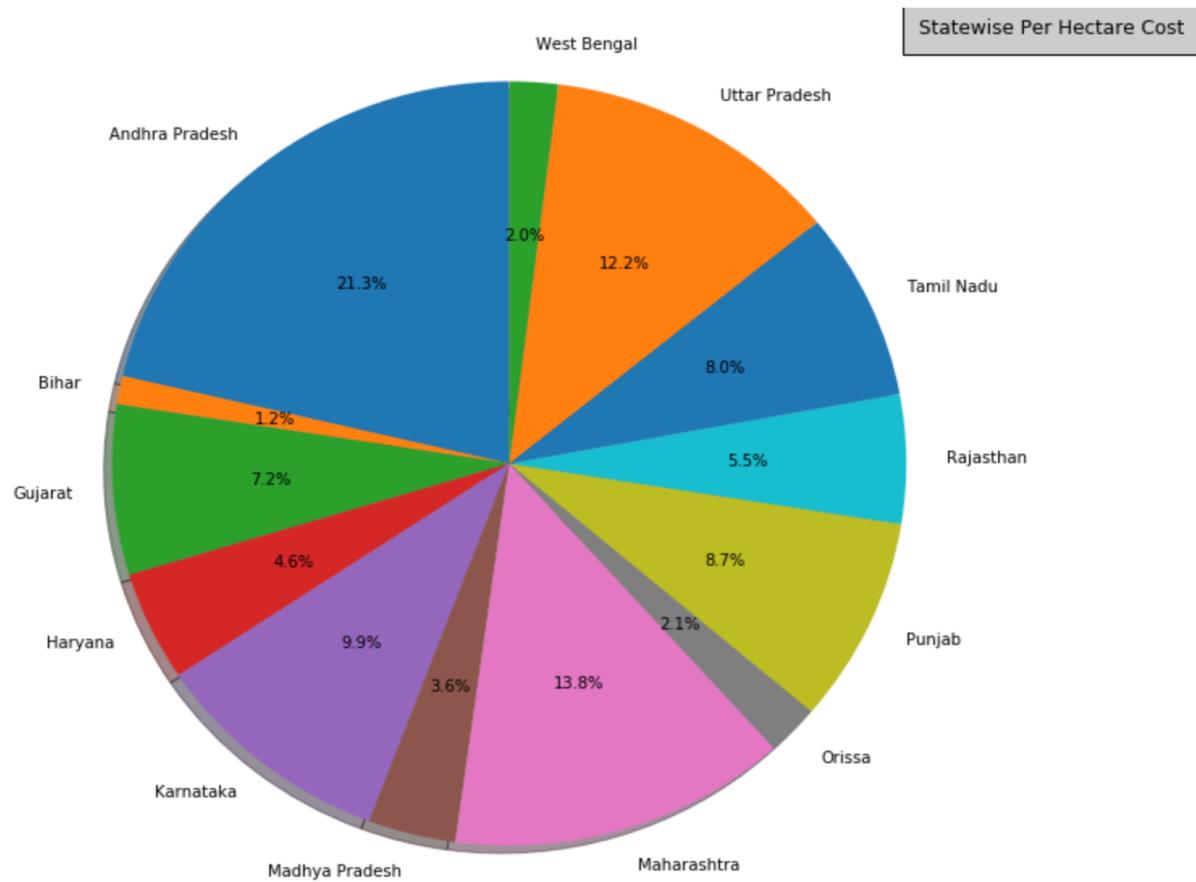


Figure 2.12: Statewise crop production, per hectare.

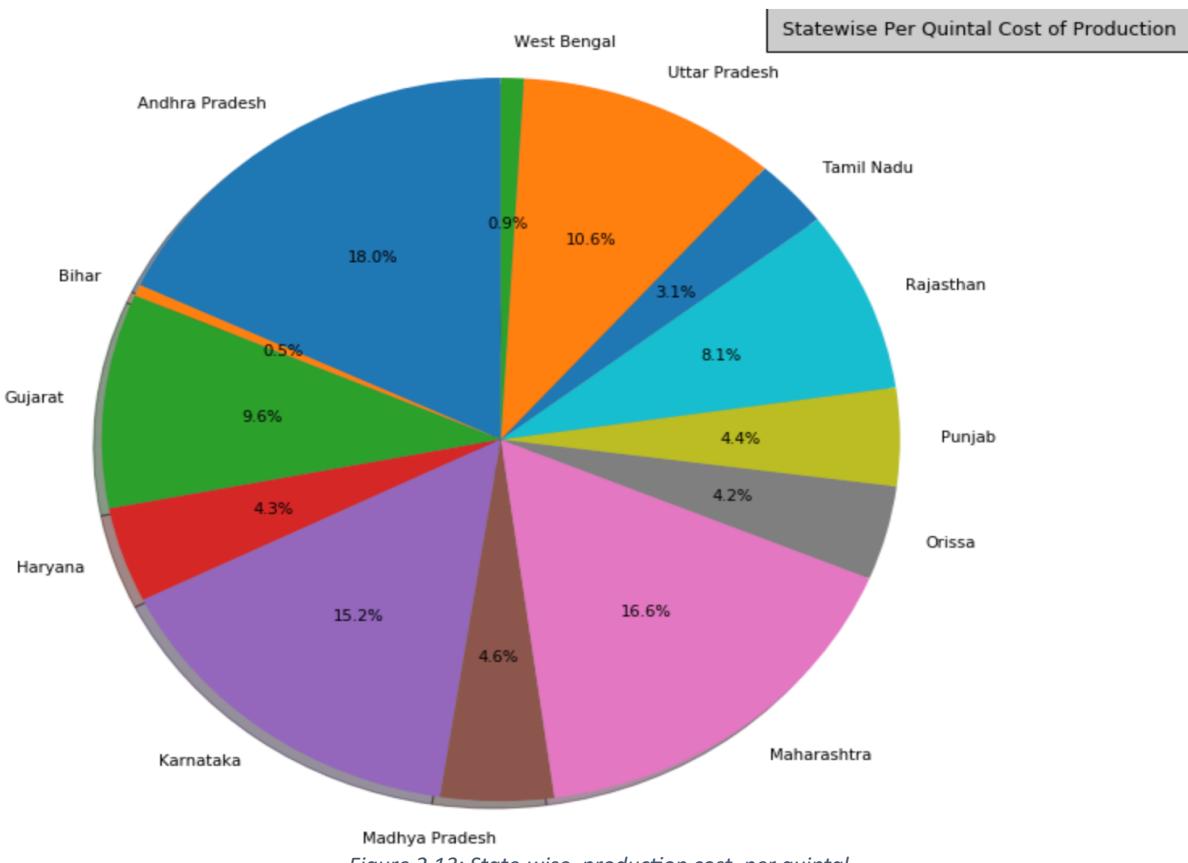


Figure 2.13: State-wise, production cost, per quintal

The need of these pie chart is to gather the information at a root level, which can help us tell someone, which crop can be an alternative source of income/cultivation for the farmers. In the current times, nobody should rely on a single source of income.

\*State-wise agricultural observations-

- Karnataka is the highest in yield.
- Odisha is least in yield.
- Andhra Pradesh and Maharashtra spend maximum in cultivation as well in production cost for agriculture.
- Bihar and West Bengal spends very less in cultivation as well in production cost for agriculture.

```
In [36]: from sklearn.ensemble import RandomForestRegressor
regr = RandomForestRegressor(max_depth=2, random_state=0, n_estimators=100)
regr.fit(a_train, b_train)
b_pred = regr.predict(a_test)

from sklearn.metrics import mean_squared_error as mse
from sklearn.metrics import mean_absolute_error as mae
from sklearn.metrics import r2_score

print('MSE =', mse(b_pred, b_test))
print('MAE =', mae(b_pred, b_test))
print('R2 Score =', r2_score(b_pred, b_test))

MSE = 7.671048879964048
MAE = 0.8953650873829122
R2 Score = 0.9589614680509005
```

Figure 2.14: Part of code showing the accuracy of the random forest model applied to the dataset for predicting the agricultural crop in some particular part of India at some point of time.

For confidential reasons, as the project is given by a particular organisation to work upon, due to confidential reasons, the code cannot be shown in entirety.

To gather the appropriate information about the figures and their usage, kindly refer to the last page, under the references column.

## **CHAPTER-3: RESULT AND CONCLUSION**

The results of this project report demonstrate the successful application of machine learning techniques in crop prediction and yield calculation for agricultural crops in India. The implemented models, including Random Forest, Neural Networks, Linear Regression, and Gradient Boosting Trees, exhibited promising performance in accurately predicting suitable crops for specific regions and estimating crop yields. The analysis revealed the importance of various features, such as historical production, weather data, soil characteristics, and cost variables, in influencing the models' predictive power. Additionally, the assessment of climate change impacts showcased the potential risks and vulnerabilities faced by different regions and crop types, emphasizing the need for proactive measures in mitigating the effects. These findings have significant implications for farmers, policymakers, and stakeholders in optimizing agricultural practices, resource allocation, and decision-making processes. By harnessing the power of machine learning, this project contributes to the advancement of data-driven approaches in agriculture, promoting sustainable and efficient crop production in India.

The results of this project report highlight the effectiveness of machine learning models in predicting crop suitability and estimating crop yields for agricultural crops in India. The models, including Random Forest, Neural Networks, Linear Regression, and Gradient Boosting Trees, exhibited robust performance and demonstrated their potential in aiding decision-making processes for farmers. The analysis emphasized the significance of diverse features, such as historical production, weather data, soil characteristics, and cost variables, in improving the accuracy of crop prediction and yield calculation models. These findings offer valuable insights into the complex dynamics between various factors influencing crop production. Furthermore, the evaluation of climate change impacts on crop yields revealed important considerations for future agricultural practices. The projected negative effects of climate change on crop yields underscore the urgency of implementing adaptive strategies and sustainable farming practices. By understanding the potential risks and vulnerabilities associated with climate change, policymakers and stakeholders can devise effective measures to mitigate its impacts and ensure food security.

The outcomes of this project provide practical implications for the agricultural sector in India. Farmers can leverage the crop prediction models to make informed decisions about crop selection, optimize resource allocation, and maximize yields. Policymakers can utilize the findings to develop policies and interventions aimed at promoting resilient agricultural practices and addressing climate change challenges. Ultimately, the integration of machine learning techniques in agriculture contributes to enhancing productivity, sustainability, and economic growth in India's agricultural sector.

In conclusion, this project demonstrates the valuable role of machine learning in crop prediction and yield estimation. The accurate prediction of crop suitability and yield calculation enables farmers and stakeholders to make informed decisions, optimize resources, and adapt to changing environmental conditions. These findings underline the potential of data-driven approaches in transforming agriculture and fostering a resilient and sustainable future for India's agricultural sector.

## **CHAPTER-4: FUTURE WORK**

In light of the results and implications of this project report, there are several potential avenues for future work to further enhance crop prediction and yield estimation in the agricultural sector. Firstly, incorporating more granular and real-time data sources, such as satellite imagery, IoT sensors, and climate models, can improve the accuracy and timeliness of predictions. By integrating these data sources, the models can capture dynamic changes in weather patterns, soil conditions, and other relevant factors, enabling more precise and up-to-date predictions. Secondly, exploring ensemble methods that combine multiple machine learning models could potentially enhance prediction performance. Ensemble techniques, such as stacking or bagging, leverage the strengths of different models to achieve more robust and accurate predictions. By blending the predictions from multiple models, the ensemble approach can mitigate individual model biases and improve overall performance.

Additionally, investigating the incorporation of domain-specific knowledge and expert systems into the models can enhance their interpretability and provide actionable insights for farmers. By integrating agricultural expertise and domain knowledge, the models can offer recommendations and strategies tailored to specific crop varieties, regional conditions, and farmer preferences. Furthermore, extending the analysis to include additional regions and crops would provide a broader understanding of crop suitability and yield estimation across different agricultural contexts. By incorporating diverse datasets from various states and crop types, the models can capture regional variations and improve their generalizability.

Lastly, considering the socio-economic aspects of agriculture, such as market prices, government policies, and farmer demographics, can enrich the models' predictive capabilities. By incorporating these factors, the models can provide insights into profitability, risk assessment, and decision-making strategies for farmers.

Overall, future work should focus on refining the existing models, incorporating additional data sources, exploring ensemble methods, integrating domain knowledge, expanding the analysis to different regions and crops, and considering socio-economic factors. By addressing these areas, the accuracy, applicability, and practicality of crop prediction and yield estimation models can be further enhanced, ultimately contributing to the advancement of agricultural practices and the sustainable development of the agricultural sector.

Something that has also been in the mind right from the very start, is making API's for the crop prediction system and weather prediction system and finally integrating them together. This can be a good model for helping the government and the farmers of our country to a great extent I personally believe.

## REFERENCES

### **About the figures:**

- 2.1- Tells us about any outlier in the dataset, which eventually helps us normalize the dataset if there's any need of it.
- 2.2- The Boxplot provides a visual summary of key statistical measures such as the median, quartiles, and outliers.
- 2.3- This plot helps us to understand about the produce of crops over various years as per the dataset.
- 2.4- This bar graph tells about the statewise production of different crops.
- 2.5- Shows the graph for better understanding of area vs yield in the initial year of the dataset, i.e., 2006-2007.
- 2.6- Shows the graph for better understanding of area vs yield in the last of the years as per the dataset, i.e., 2010-2011.
- 2.7- Shows a comparative plot between the production in year 2006-2007 and 2010-2011.
- 2.8- Shows a comparative study between the area in yield between 2006-2007 and 2010-2011.
- 2.9- Shows the comparative plot about the final yield between 2006-2007 and 2010-2011.

### **Journal/Research Paper references:**

1. Ahamed, A.T.M.S., Mahmood, N.T., Hossain, N., Kabir, M.T., Das, K., Rahman, F., Rahman, R.M., 2015. Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh. In: 2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD 2015 - Proceedings, <https://doi.org/10.1109/SNPD.2015.7176185>.
2. Ahmad, I., Saeed, U., Fahad, M., Ullah, A., Habib-ur-Rahman, M., Ahmad, A., Judge, J., 2018. Yield forecasting of spring maize using remote sensing and crop modeling in Faisalabad-Punjab Pakistan. *J. Indian Soc. Remote Sens.* 46 (10), 1701–1711. <https://doi.org/10.1007/s12524-018-0825-8>.
3. Ali, I., Cawkwell, F., Dwyer, E., Green, S., 2017. Modeling managed grassland biomass estimation by using multitemporal remote sensing data—a machine learning approach. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 10 (7), 3254–3264. <https://doi.org/10.1109/JSTARS.2016.2561618>.
4. Alpaydin, E., 2010. Introduction to Machine Learning, 2nd ed. Retrieved from [https://books.google.nl/books?hl=nl&lr=&id=TtrxCwAAQBAJ&oi=fnd&pg=PR7&dq=introduction+to+machine+learning&ots=T5ejQG\\_7pZ&sig=0xC\\_H0agN7mPhYW7oQsWiMVwRnQ#v=onepage&q=introduction to machine learning&f=false](https://books.google.nl/books?hl=nl&lr=&id=TtrxCwAAQBAJ&oi=fnd&pg=PR7&dq=introduction+to+machine+learning&ots=T5ejQG_7pZ&sig=0xC_H0agN7mPhYW7oQsWiMVwRnQ#v=onepage&q=introduction to machine learning&f=false).
5. Ananthara, M.G., Arunkumar, T., Hemavathy, R., 2013. CRY-An improved crop yield prediction model using bee hive clustering approach for agricultural data sets. In: Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering, PRIME 2013, 473–478. <https://doi.org/10.1109/ICPRIME.2013.6496717>.
6. Ayodele, T.O., 2010. Introduction to Machine Learning.
7. Baldi, P., 2012. Autoencoders, unsupervised learning, and deep architectures. In: Proceedings of ICML workshop on unsupervised and transfer learning, pp. 37–49.
8. Baral, S., Kumar Tripathy, A., Bijayasingh, P., 2011. Yield Prediction Using Artificial Neural Networks, pp. 315–317. [https://doi.org/10.1007/978-3-642-19542-6\\_57](https://doi.org/10.1007/978-3-642-19542-6_57).
9. Bargoti, S., Underwood, J.P., 2017. Image segmentation for fruit detection and yield estimation in apple orchards. *J. Field Rob.* 34 (6), 1039–1060. <https://doi.org/10.1002/rob.21699>.

10. Gandhi, N., Petkar, O., Armstrong, L.J., Tripathy, A.K., 2016. Rice crop yield prediction in India using support vector machines. In: 2016 13th International Joint Conference on Computer Science and Software Engineering, JCSSE 2016. <https://doi.org/10.1109/JCSSE.2016.7748856>.
11. Girish, L., Gangadhar, S., Bharath, T., Balaji, K., n.d. Crop Yield and Rainfall Prediction in Tumakuru District using Machine Learning. Ijream.Org. Retrieved from <https://www.ijream.org/papers/NCTFRD2018015.pdf>.