SRI RAMACHANDRA
INSTITUTE OF HIGHER EDUCATION AND RESEARCH
(Category - I Deemed to be University) Porur, Chennai
SRI RAMACHANDRA FACULTY OF ENGINEERING AND TECHNOLOGY

Sustainable and Efficient Crop Production: Leveraging Data Analytics to Address the Challenges of Climate Change

INT 500 – INTERNSHIP 4 - PROJECT REPORT

Submitted by

YESWANTH R – E0119031

In partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

(Artificial Intelligence and Machine Learning)

Sri Ramachandra Faculty of Engineering and Technology

Sri Ramachandra Institute of Higher Education and Research, Porur, Chennai - 600116

April, 2023

Sustainable and Efficient Crop Production: Leveraging Data Analytics to Address the Challenges of Climate Change

INT 500 – INTERNSHIP 4 - PROJECT REPORT

Submitted by

YESWANTH R – E0119031

In partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE AND ENGINEERING

(Artificial Intelligence and Machine Learning)

Sri Ramachandra Faculty of Engineering and Technology

Sri Ramachandra Institute of Higher Education and Research, Porur, Chennai - 600116

April, 2023

SRI RAMACHANDRA
INSTITUTE OF HIGHER EDUCATION AND RESEARCH
(Category - I Deemed to be University) Porur, Chennai
SRI RAMACHANDRA FACULTY OF ENGINEERING AND TECHNOLOGY

BONAFIDE CERTIFICATE

Certified that this project report **"Sustainable and Efficient Crop Production: Leveraging Data Analytics to Address the Challenges of Climate Change"** is the bonafide record of work done by **"YESWANTH R – E0119031"** who carried out the internship work under my supervision.

Signature of the Supervisor

Signature of Programme Coordinator

Dr. G. Shiyamala Gowri

Assistant Professor,

Department of Artificial Intelligence and Data Analytics

Sri Ramachandra Faculty of Engineering and Technology,

SRIHER, Porur, Chennai-600 116.

Prof. Vanitha Magesh

Assistant Professor,

Department of Artificial Intelligence and Machine Learning

Sri Ramachandra Faculty of Engineering and Technology,

SRIHER, Porur, Chennai-600 116.

Evaluation Date:

ACKNOWLEDGEMENT

I express my sincere gratitude to our Programme Coordinator **Prof. Vanitha Magesh** for their support and for providing the required facilities for carrying out this study.

I wish to thank my faculty supervisor(s), **Prof. Shiyamala Gowri**, Department of Artificial Intelligence and Data Analytics, Sri Ramachandra faculty of Engineering and Technology for extending help and encouragement throughout the project. Without her continuous guidance and persistent help, this project would not have been a success for me.

I am grateful to all the members of Sri Ramachandra Faculty of Engineering and Technology, my beloved parents and friends for extending the support, who helped us to overcome obstacles in the study.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

This abstract describes the use of data analytics in sustainable and efficient crop production to tackle the challenges of climate change. By leveraging data-driven insights, farmers can make informed decisions to optimize resource use and improve crop yields, while reducing environmental impact. This approach can contribute to a more resilient and sustainable food system.

As the global population continues to grow, the need for sustainable and efficient crop production becomes increasingly important, particularly in the face of climate change.

This report explores how data analytics can be leveraged to address the challenges of climate change in crop production, including the use of precision agriculture, remote sensing, and machine learning techniques.

Through the application of data analytics in agriculture, sustainable and efficient crop production practices can be implemented to mitigate the impact of climate change on food security.

CHAPTER - 1

INTRODUCTION

Digital agriculture refers to the integration of technology, data, and communication tools into traditional agricultural practices. It involves the use of sensors, drones, artificial intelligence, machine learning, and other digital tools to optimize and streamline agricultural processes.

The goal of digital agriculture is to improve crop yields, reduce waste, and increase efficiency, all while minimizing the environmental impact of farming.

The agriculture sector in India is vulnerable to climate change. Higher temperatures tend to reduce crop yields and favour weed and pest proliferation.

Climate change can negatively affect irrigated crop yields across agroecological regions due to temperature rise and changes in water availability.

Taking the various parameters for a crop to grow well and increase productivity and finding out the best crop to grow can be done for any area with the right variables.

This project focuses on the efficiency and maximization of yield produced considering various climatic variables using AIML and Data analytics.

CHAPTER - 2

LITERATURE SURVEY

2.1 Research Paper 1

Title: “Analysis of agriculture data using data mining techniques: application of big data”
Authors: Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki
Published: 2017.
In agriculture sector where farmers and agribusinesses have to make innumerable decisions every day and intricate complexities involves the various factors influencing them. An essential issue for agricultural planning intention is the accurate yield estimation for the numerous crops involved in the planning. Data mining techniques are necessary approach for accomplishing practical and effective solutions for this problem. Agriculture has been an obvious target for big data. Environmental conditions, variability in soil, input levels, combinations and commodity prices have made it all the more relevant for farmers to use information and get help to make critical farming decisions. This paper focuses on the analysis of the agriculture data and finding optimal parameters to maximize the crop production using data mining techniques like PAM, CLARA, DBSCAN and Multiple Linear Regression. Mining the large amount of existing crop, soil and climatic data, and analysing new, non-experimental data optimizes the production and makes agriculture more resilient to climatic change.

2.2 Research paper 2

Title: “Research Paper on Big Data Analytics for Agricultural Development in India”
Authors: Saumya Pathak, Manjeet Kumar Singh, Sushant Shekhar Gautam.
Published: IJRECE VOL. 7 ISSUE 2 APR.-JUNE 2019
In India, dominant parts of provincial inhabitants are reliant on agriculture for their occupation. Be that as it may, the current farming practices are not prudentially suitable neither naturally supportable and the yields of numerous agrarian items in India are basically low. Sooner rather than later, it will get to be the key for the nation to fabricate a high yielding, focused, and shifted agricultural part and speed up country, non-ranch business enterprise.
This paper recognizes the culminations of conventional cultivating practices and delivers how to build the yield of the agrarian things by utilizing present day PC innovations. Further, it additionally recognizes the basic figuring and analytic capacity of Big Data in preparing colossal volumes of value-based information continuously circumstances. The target of this paper is to display the revisions in the rural area and supports the dialogs on how government can cultivate developments in the enormous information examination to enhance the rustic agricultural framework.

2.3 Research paper 3

Title: "Data analytics for crop management: a big data view"
Authors: Nabila Chergui and Mohand Tahar Kechadi.
Published: 2022
Recent advances in Information and Communication Technologies have a significant impact on all sectors of the economy worldwide. Digital Agriculture appeared as a consequence of the democratisation of digital devices and advances in artificial intelligence and data science. Digital agriculture created new processes for making farming more productive and efficient while respecting the environment. Recent and sophisticated digital devices and data science allowed the collection and analysis of vast amounts of agricultural datasets to help farmers, agronomists, and professionals understand better farming tasks and make better decisions. In this paper, we present a systematic review of the application of data mining techniques to digital agriculture. We introduce the crop yield management process and its components while limiting this study to crop yield and monitoring. After identifying the main categories of data mining techniques for crop yield monitoring, we discuss a panoply of existing works on the use of data analytics. This is followed by a general analysis and discussion on the impact of big data on agriculture.

2.4 Research paper 4

Title: "Building capacity for assessing spatial-based sustainability metrics in agriculture"
Authors: Louis Kouadio and Nathaniel K Newlands.
Published: 2015
Crop yield is influenced over time and space, namely, by a wide range of variables linked with crop genetics, agronomic management practices and the environment under which the crop dynamically responds to maximize growth potential and survival. Here, decision analytics can play a vital role by guiding the use of statistical-based analytics to build in a higher degree of intelligence to enable better predictive and prescriptive approaches. While inter-annual variability in yield can be modelled based on a deterministic trend with stochastic variation, quantifying the variability of yield and how it changes across different spatial resolutions remains a major knowledge gap. To better understand how yield scales spatially, we integrate in this study, for the first time, multi-scale crop yield of spring wheat and its variance obtained within the major wheat growing region of the Canadian Prairies. We found large differences between the mean and variance from field to district to regional scales, from which we determined spatially-dependent scaling factors for the mean and variance of crop yield. From our analysis, we provide several key recommendations for building capacity in assessing agricultural sustainability using spatial-based metrics. In the future, the use of such metrics may broaden the adoption and consistent implementation of new sustainable management protocols and practices under a precautionary, adaptive management approach.

CHAPTER – 3

TOOLS AND TECHNIQUES

3.1 Google Colab

Colab is a free Jupyter notebook environment that runs entirely in the cloud. Most importantly, it does not require a setup and the notebooks that you create can be simultaneously edited by your team members - just the way you edit documents in Google Docs. Colab supports many popular machine learning libraries.

3.2 Python

Python is a high-level, general-purpose programming language. Its design philosophy emphasizes code readability with the use of significant indentation via the off-side rule. Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including structured, object-oriented and functional programming.

3.3 R Studio

RStudio is an integrated development environment for R, a programming language for statistical computing and graphics. It is available in two formats: RStudio Desktop is a regular desktop application while RStudio Server runs on a remote server and allows accessing RStudio using a web browser.

3.4 Highcharts

Highcharts is a pure JavaScript based charting library meant to enhance web applications by adding interactive charting capability. Highcharts provides a wide variety of charts. For example, line charts, spline charts, area charts, bar charts, pie charts and so on.

CHAPTER – 4

METHODOLOGY / WORKFLOW

Though there are many yield prediction models, they are neither fully functional nor implemented fully in the real world. So, we have thought to make our proposed system fully functional and also develop in a simple manner.

4.1 System Architecture

The below diagram depicts the system architecture of our project. Our whole system can be divided into 2 modules i.e., one model predicts the optimal yield and the other model analyses the patterns in the dataset. The operation of these models is specified clearly in the above diagram.

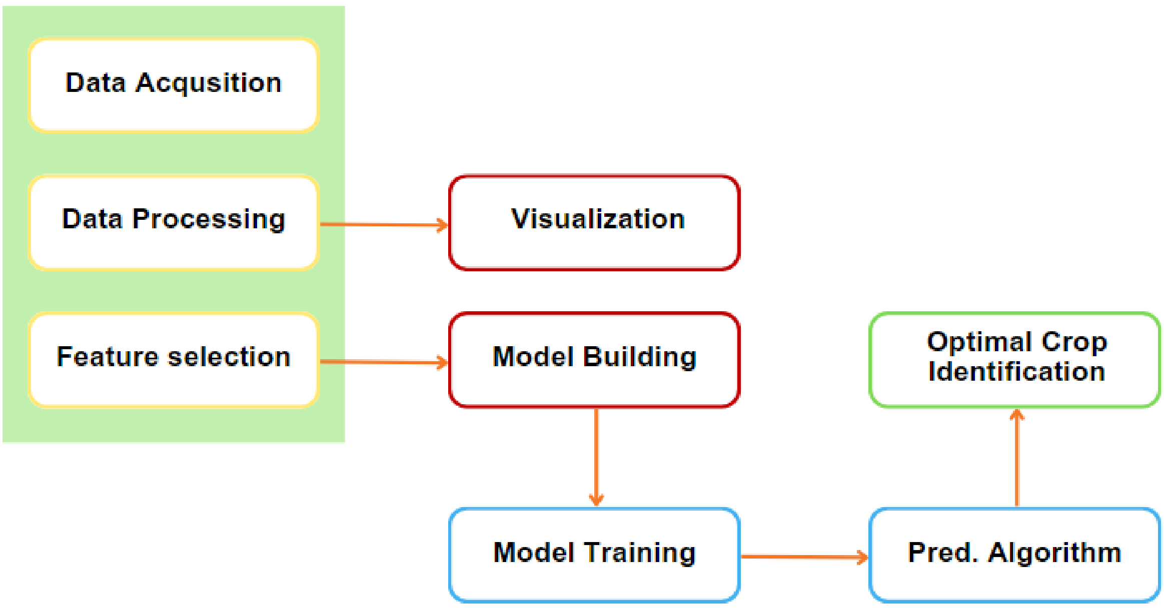


Fig 4.1.1 Model workflow

CHAPTER – 5

RESULTS AND DISCUSSION

5.1 Dataset information / Pre-Processing

The input dataset consists of data with the following parameters namely: Crop (cotton, groundnut, jowar, rice, and some fruit crops, etc.), Temperature, average rainfall (mm), soil, PH value, soil type, major fertilizers, nitrogen (kg/Ha), phosphorus (Kg/Ha), Potassium (Kg/Ha), minimum rainfall required, the minimum temperature required.

	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

Fig 5.1.1 Input Data

In data pre-processing, we check for any null values and remove them as they may affect efficiency and get info about the dataset.

```
crop.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2200 entries, 0 to 2199
Data columns (total 8 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   N            2200 non-null   int64
 1   P            2200 non-null   int64
 2   K            2200 non-null   int64
 3   temperature  2200 non-null   float64
 4   humidity     2200 non-null   float64
 5   ph           2200 non-null   float64
 6   rainfall     2200 non-null   float64
 7   label        2200 non-null   object
dtypes: float64(4), int64(3), object(1)
memory usage: 137.6+ KB
```

```
crop.drop_duplicates()
```

	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

Fig 5.1.2 Pre-Processing

5.2 Summary Statistics

A statistical summary can help identify outliers, missing values, and other anomalies in the dataset, and provide insights into the overall patterns and trends of the data. It can also be used to compare different datasets and evaluate the effectiveness of data processing and analysis techniques.

	N	P	K	temperature	humidity	ph	rainfall	label
count	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000	2200.000000	2200
unique	NaN	NaN	NaN	NaN	NaN	NaN	NaN	22
top	NaN	NaN	NaN	NaN	NaN	NaN	NaN	rice
freq	NaN	NaN	NaN	NaN	NaN	NaN	NaN	100
mean	50.551818	53.362727	48.149091	25.616244	71.481779	6.469480	103.463655	NaN
std	36.917334	32.985883	50.647931	5.063749	22.263812	0.773938	54.958389	NaN
min	0.000000	5.000000	5.000000	8.825675	14.258040	3.504752	20.211267	NaN
25%	21.000000	28.000000	20.000000	22.769375	60.261953	5.971693	64.551686	NaN
50%	37.000000	51.000000	32.000000	25.598693	80.473146	6.425045	94.867624	NaN
75%	84.250000	68.000000	49.000000	28.561654	89.948771	6.923643	124.267508	NaN
max	140.000000	145.000000	205.000000	43.675493	99.981876	9.935091	298.560117	NaN

Fig 5.2.1 Statistical summary of data

Displaying the average values of all the variables for each crop available in the dataset.

```
conditions  N
Average Value for N is 50.55
----------------------------------------
Rice : 79.89
Black Grams : 40.02
Banana : 100.23
Jute : 78.40
Coconut : 21.98
Apple : 20.80
Papaya : 49.88
Muskmelon : 100.32
Grapes : 23.18
Watermelon : 99.42
Kidney Beans: 20.75
Mung Beans : 20.99
Oranges : 19.58
Chick Peas : 40.09
Lentils : 18.77
Cotton : 117.77
Maize : 77.76
Moth Beans : 21.44
Pigeon Peas : 20.73
Mango : 20.07
Pomegranate : 18.87
Coffee : 101.20
```

Fig 5.2.2 Average variable values

5.3 Feature Selection

In this step, we process available features and select the final parameters that are influential in nature. The list of final parameters is given below.

Crop	Name of the crop
pH Level	This describes the nature of the soil
Nitrogen	Amount of nitrogen present
Potassium	Amount of potassium present
Phosphorus	Amount of phosphorus present
Rainfall	Expected rainfall in millimeters
Temperature	Optimal temperature for the crop
Humidity	Water Vapour content available for the crop

Fig 5.3.1 Important Features

5.4 Accuracy Comparison

Here in this project four classification algorithms like **Decision tree, Random Forest, Logistic Regression and Gradient Boosting** were implemented and their accuracies is compared to get better prediction to find the optimal crop.

Decision Tree

```
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import cross_val_score
import sklearn.metrics as metrics

DecisionTree = DecisionTreeClassifier
(criterion="entropy",random_state=2,max_depth=5)
DecisionTree.fit(Xtrain,Ytrain)

predicted_values = DecisionTree.predict(Xtest)
x = metrics.accuracy_score(Ytest, predicted_values)
acc.append(x)
model.append('Decision Tree')
print("DecisionTrees's Accuracy is: ", x*100)

print(classification_report(Ytest,predicted_values))
```

```
DecisionTrees's Accuracy is:  90.0
```

Fig 5.4.1 DT model

Random Forest

```
RF = RandomForestClassifier(n_estimators=20, random_state=0)
RF.fit(Xtrain,Ytrain)

predicted_values = RF.predict(Xtest)

x = metrics.accuracy_score(Ytest, predicted_values)
acc.append(x)
model.append('RF')
print("RF's Accuracy is: ", x*100)

print(classification_report(Ytest,predicted_values))
```

```
RF's Accuracy is:  99.0909090909091
```

Fig 5.4.2 RF model

Logistic Regression

```python
from sklearn.linear_model import LogisticRegression

LogReg = LogisticRegression(random_state=2)

LogReg.fit(Xtrain,Ytrain)

predicted_values = LogReg.predict(Xtest)

x = metrics.accuracy_score(Ytest, predicted_values)
acc.append(x)
model.append('Logistic Regression')
print("Logistic Regression's Accuracy is: ", x*100)

print(classification_report(Ytest,predicted_values))
```

```
Logistic Regression's Accuracy is:  95.22727272727273
```

Fig 5.4.3 LR model

Gradient Boosting

```python
from sklearn.ensemble import GradientBoostingClassifier
grad = GradientBoostingClassifier().fit(Xtrain, Ytrain)

predicted_values = grad.predict(Xtest)
x = metrics.accuracy_score(Ytest, predicted_values)
acc.append(x)
model.append('Gradient Boosting')
print("Gradient Boosting Accuracy is: ", x*100)

print(classification_report(Ytest,predicted_values))
```

```
Gradient Boosting Accuracy is:  99.31818181818181
```

Fig 5.4.4 GB model

5.5 Evaluation Metrics

Cross-validation may be a statistical procedure that is used to estimate the skill of machine learning models it gives a more accurate measure of model quality. Cross-validation results of Random Forest, Decision Tree, Logistic regression, and Gradient Boosting.

```python
# Cross validation score (Decision Tree)
score = cross_val_score(DecisionTree, features, target,cv=5)
score
```

```
array([0.93636364, 0.90909091, 0.91818182, 0.87045455, 0.93636364])
```

```python
# Cross validation score (Random Forest)
score = cross_val_score(RF,features,target,cv=5)
score
```

```
array([0.99772727, 0.99545455, 0.99772727, 0.99318182, 0.98863636])
```

```python
# Cross validation score (Logistic Regression)
score = cross_val_score(LogReg,features,target,cv=5)
score
```

```
array([0.95      , 0.96590909, 0.94772727, 0.96590909, 0.94318182])
```

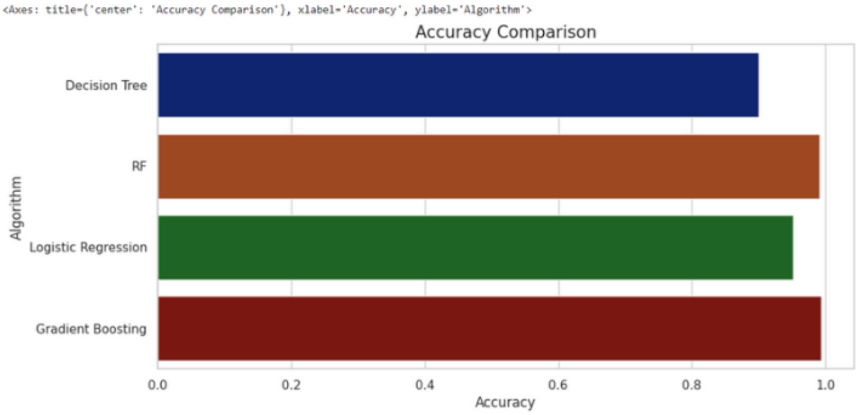
```python
# Cross validation score (Gradient Boosting)
score = cross_val_score(grad,features,target,cv=5)
score
```

```
array([0.99090909, 0.98863636, 0.99318182, 0.99772727, 0.98636364])
```

Fig 5.5.1 Cross-validation scores

Here we can see that all models scores were above 90 which indicates good model and the predictions would be accurate.

```
<Axes: title={'center': 'Accuracy Comparison'}, xlabel='Accuracy', ylabel='Algorithm'>
```



	Model	Accuracy
0	Decision Tree	0.900000
1	RF	0.990909
2	Logistic Regression	0.952273
3	Gradient Boosting	0.993182

Fig 5.5.2 Accuracy Comparison

Here comparing all four models and their performance metrics we plot the accuracy of each model. According to the analyses of cross-evaluation metrics, all four algorithms work well, but gradient boosting gives a slightly better accuracy score on test data than the other four models.

5.6 Model Prediction

The built machine learning model is applied to new data to make predictions and obtain results it is a crucial component of predictive analytics, a type of data analytics that uses current and historical data to forecast activity, behaviour, and trends.

	Real_class	Predicted_class
2121	coffee	coffee
960	pomegranate	pomegranate
952	pomegranate	pomegranate
1958	cotton	cotton
681	mungbean	mungbean

Fig 5.6.1 Obtained predictions

This is the obtained dataframe having predicted class values after passing through the model. As the Gradient boosting and random forest having higher accuracy these two models were taken to predict the optimal crop by passing the variable values.

Optimal Crop for given climatic variables

```
[ ] data = np.array([[34,60,22,17.66148158,18.15302753,5.635231778,100.6711761]])
    prediction = RF.predict(data)
    print("Given Parameters best suit's for the crop : ",prediction)

    Given Parameters best suit's for the crop :  ['kidneybeans']

[ ] data = np.array([[37,62,22,29.03,64.49,7.47,54.93]])
    prediction = grad.predict(data)
    print("Given Parameters best suit's for the crop : ",prediction)

    Given Parameters best suit's for the crop :  ['lentil']
```

Fig 5.6.2 Optimal crop recommendation

CHAPTER – 6

VISUALIZATION INSIGHTS

6.1 Distribution of Agricultural Variables

Comparison between Nitrogen, Potassium, Phosphorous values for each crop and we can see that apple and grapes need higher potash and phosphorous comparing the rest. Some crops need higher nitrogen but mostly nitrogen is average for other crops.

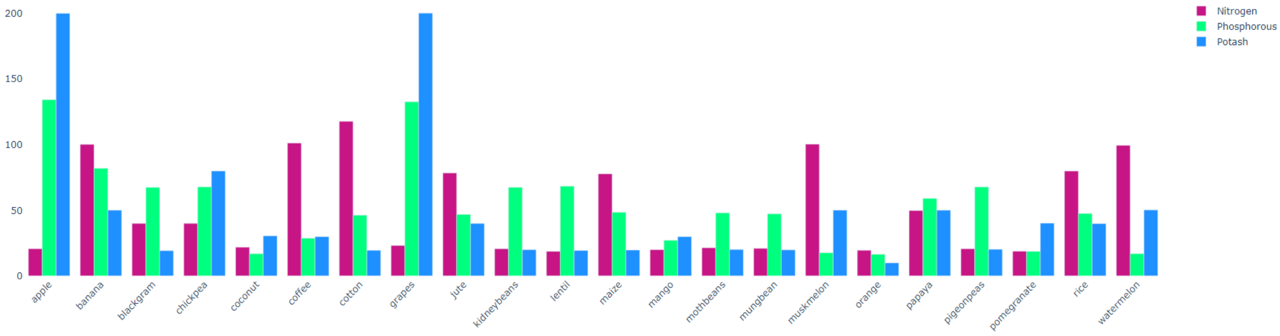


Fig 6.1.1 NPK comparison

Distribution of all features that are needed for a crop is plotted using distplot.

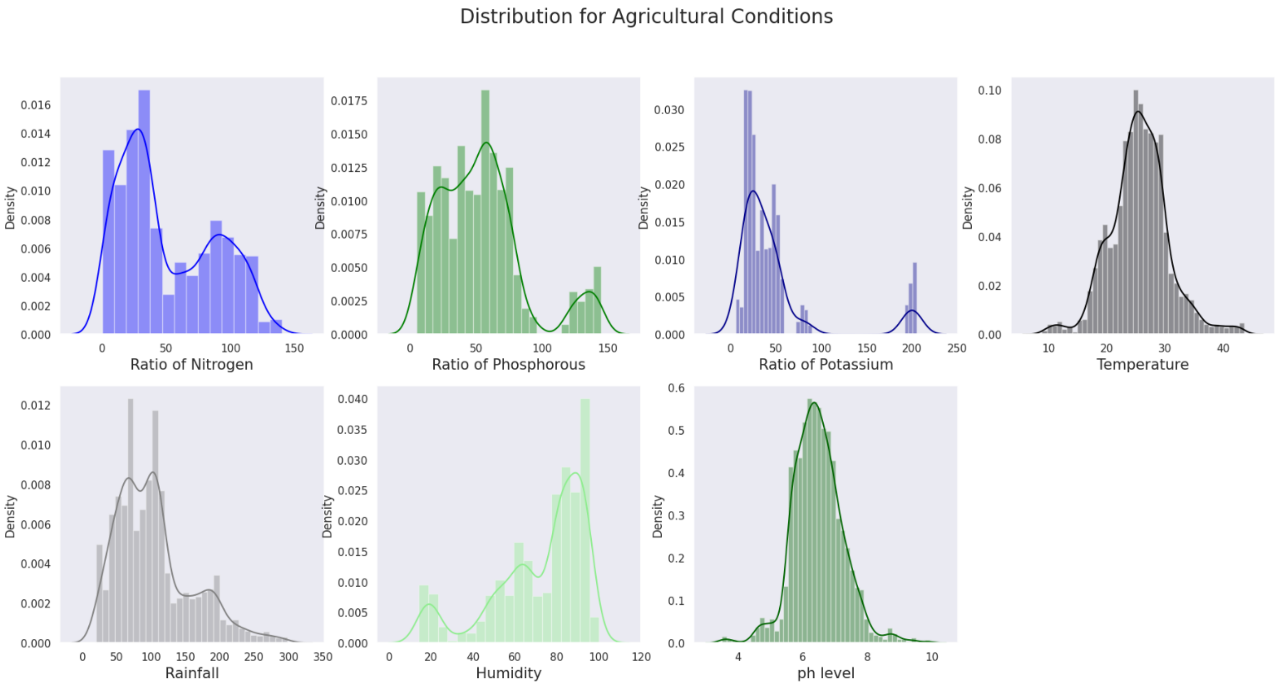


Fig 6.1.2 Distribution of Features

Correlation plot between all the features and we can see there is a good correlation between Phosphorous and Potassium.

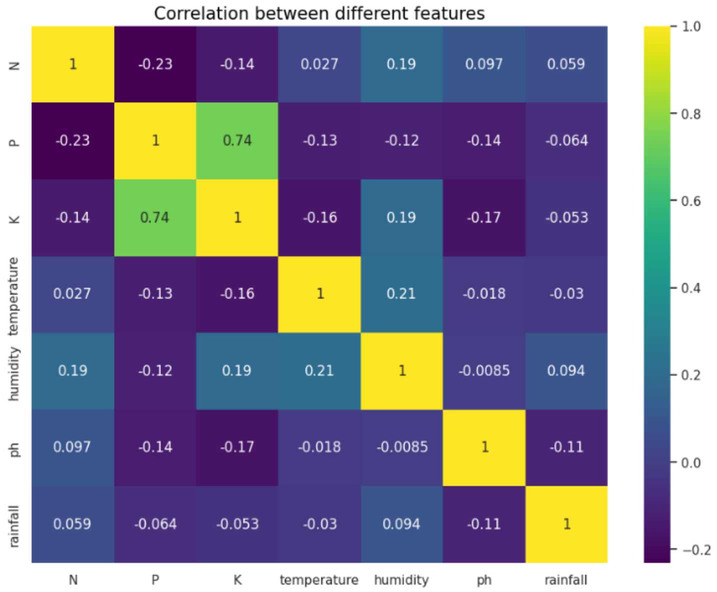


Fig 6.1.3 Correlation Plot

6.2 Rainfall and Nutrition

The first Jointplot is plotted between variables rainfall, humidity. During rainy season, average rainfall is high (average 120 mm) and temperature is mildly chill (less than 30'C). Rain affects soil moisture which affects Ph of the soil. Here are the crops which are likely to be planted during this season. Rice needs heavy rainfall (>200 mm) and a humidity above 80%. No wonder major rice production in India comes from East Coasts which has average of 220 mm rainfall every year.

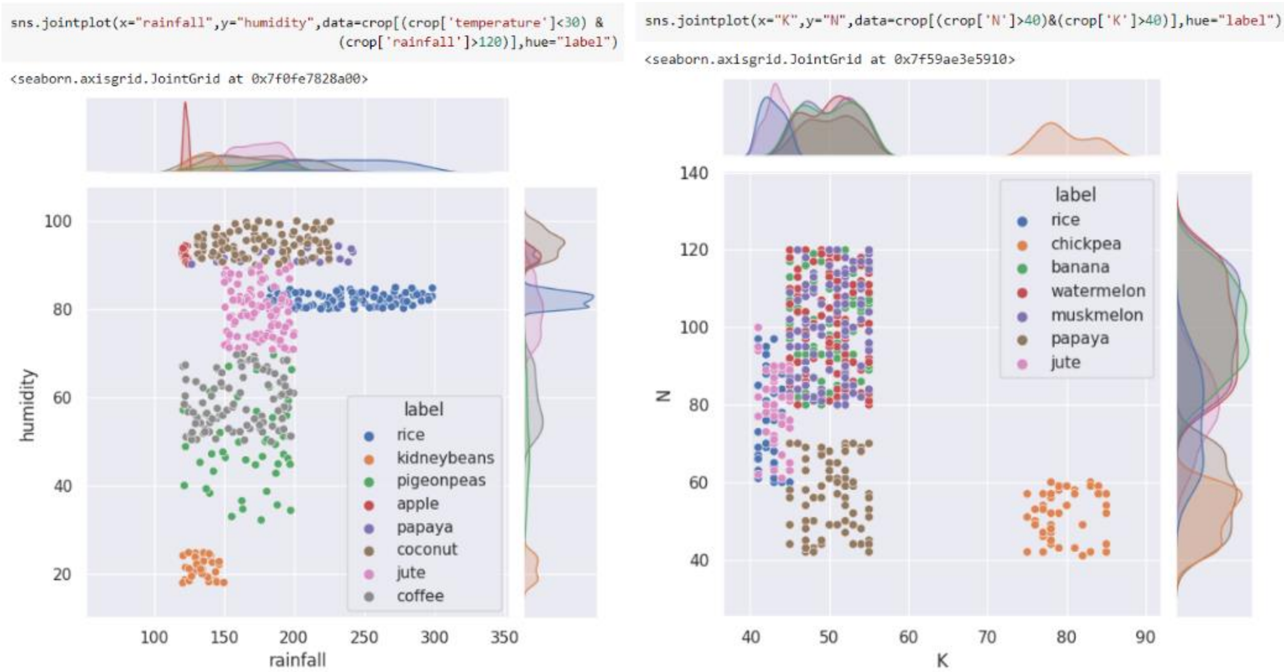


Fig 6.2.1 Comparing rainfall, humidity, K and N

The second graph which correlates with average potassium (K) and average nitrogen (N) value (both>50). These soil ingredients directly affect nutrition value of the food. Fruits which have high nutrients typically has consistent potassium values.

6.3 Optimal Temperature and Ph

The first Jointplot graph correlates with temperature and humidity values. Here the optimal temperature and humidity ranges from 10-40 and 40-100 for most of the crops respectively. From the second graph we can see Ph values are critical when it comes to soil. A stability between 6 and 7 is preferred for most of the crops.

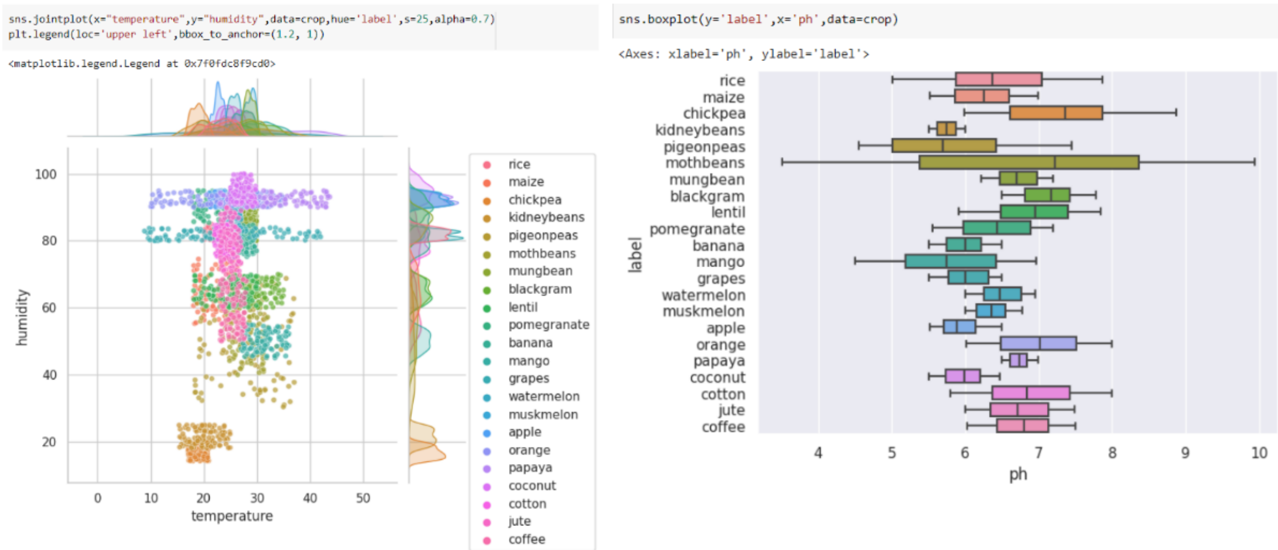


Fig 6.3.1 Optimal Temp and Ph

6.4 Analysing Phosphorus

Further analysing phosphorous levels. When humidity is less than 65, almost same Phosphor levels (approx.14 to 25) are required for 6 crops which could be grown just based on the amount of rain expected over the next few weeks. Another interesting analysis where Phosphorous levels are quite differentiable when it rains heavily (above 150 mm).

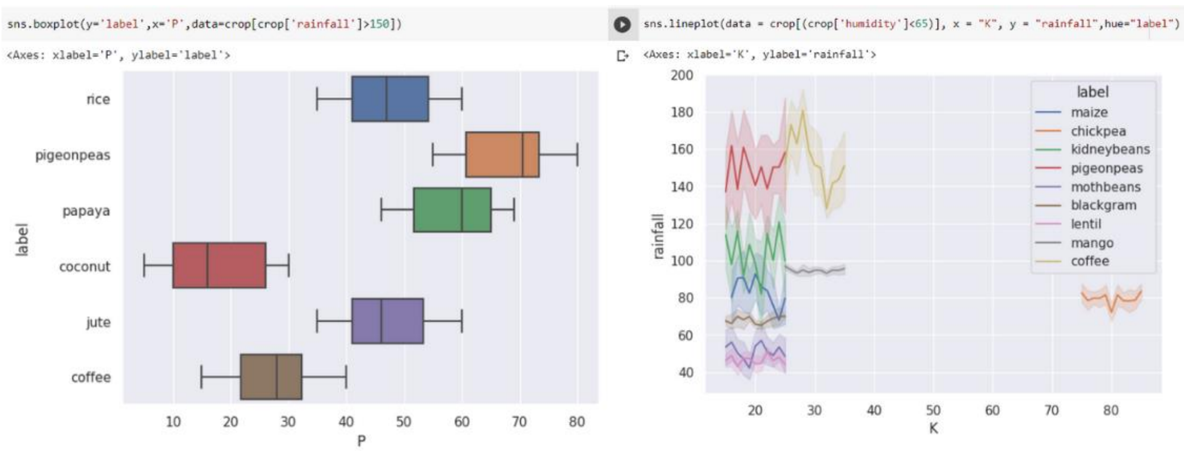


Fig 6.4.1 Phosphorous Analysis

6.5 Country's Crop Production

The Highcharter library is used to visualize the state-wise production from the year 1997 to 2021 and their yield. Here tamilnadu's production is highlighted and pie-chart of season wise production is plotted and we can see that kharif and rabi are playing a major role in production.

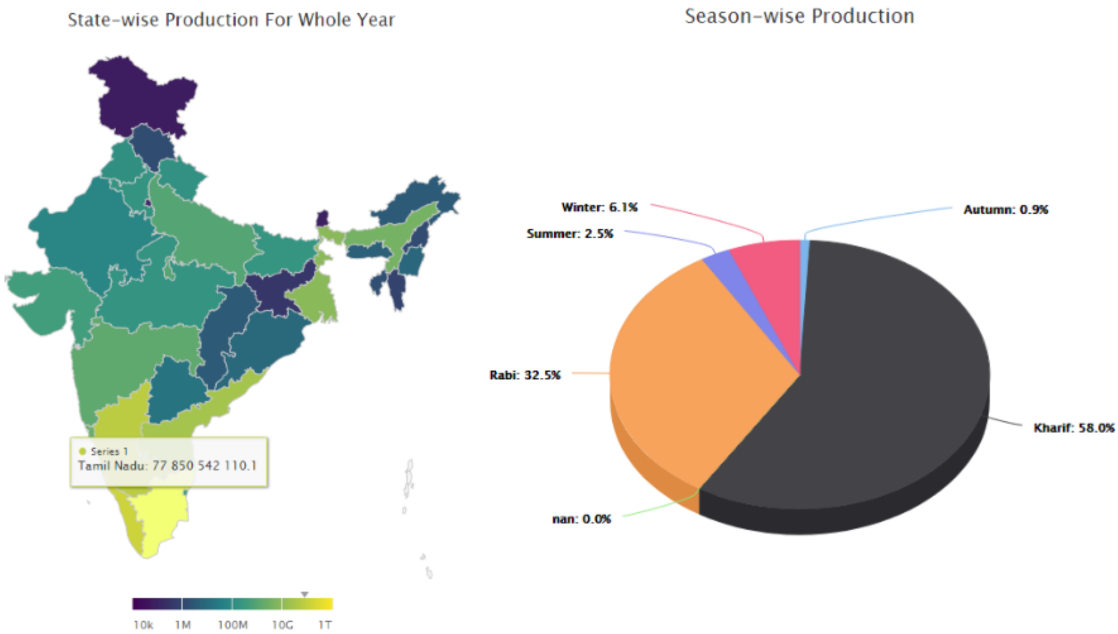


Fig 6.5.1 State and season production

6.6 Country's Season-wise Production

All five seasons production all across each state were visualized over the years and we can see most production happens in kharif, rabi and summer seasons.

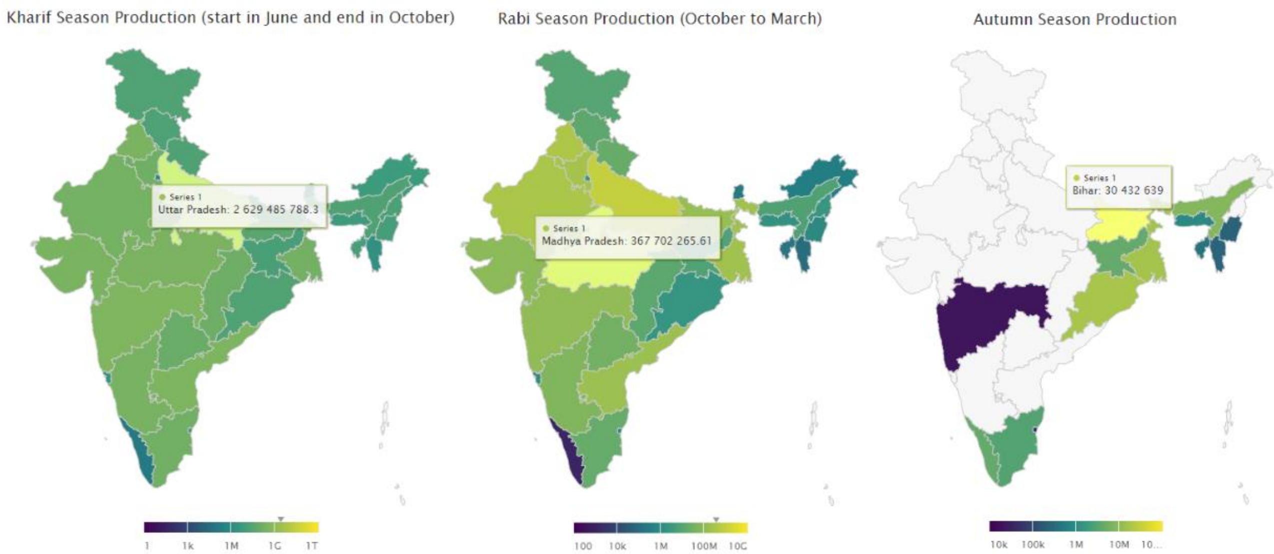


Fig 6.6.1 Kharif, Rabi and Autumn Production

Winter and autumn season is where the production is minimal over the years and top ten years of production is plotted to get insights about the past crop yields.

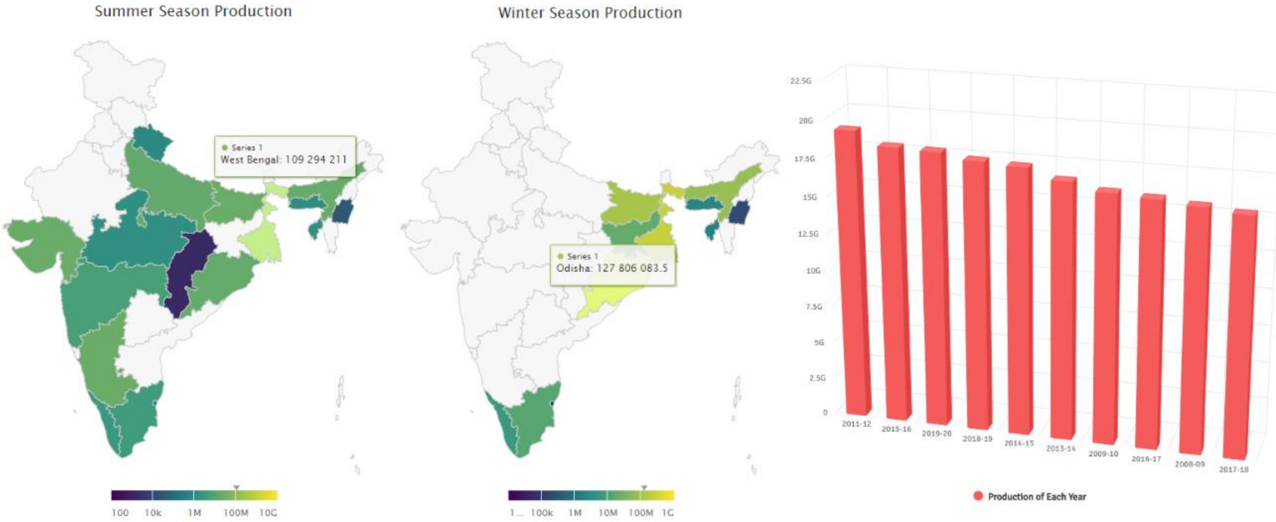


Fig 6.6.2 Summer, Winter Production

6.7 State-wise Highest produced Crop

Plotting the state-wise highly produced crops and their yield respectively here we can see the highly produced crop in Rajasthan is Wheat.

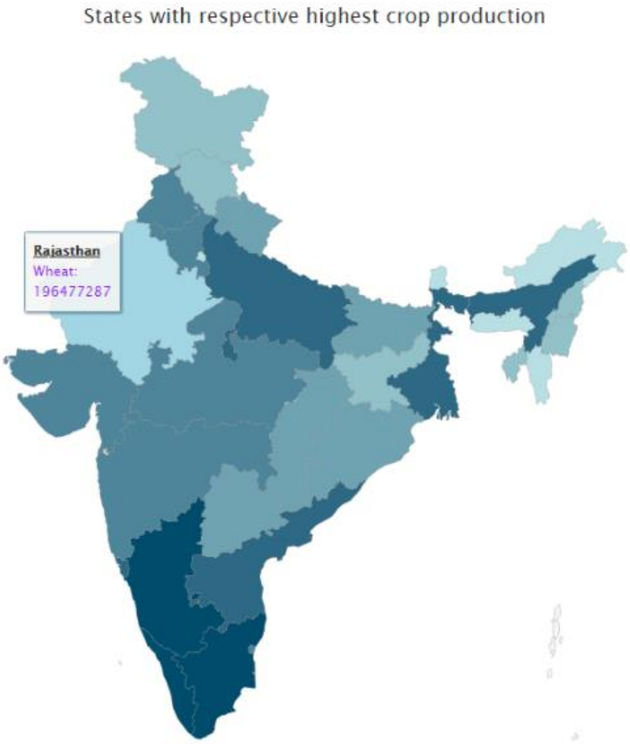


Fig 6.7.1 Highly produced crop

Line chart to visualize the season-wise crop production across all five seasons and over the years from 1997 to 2021.

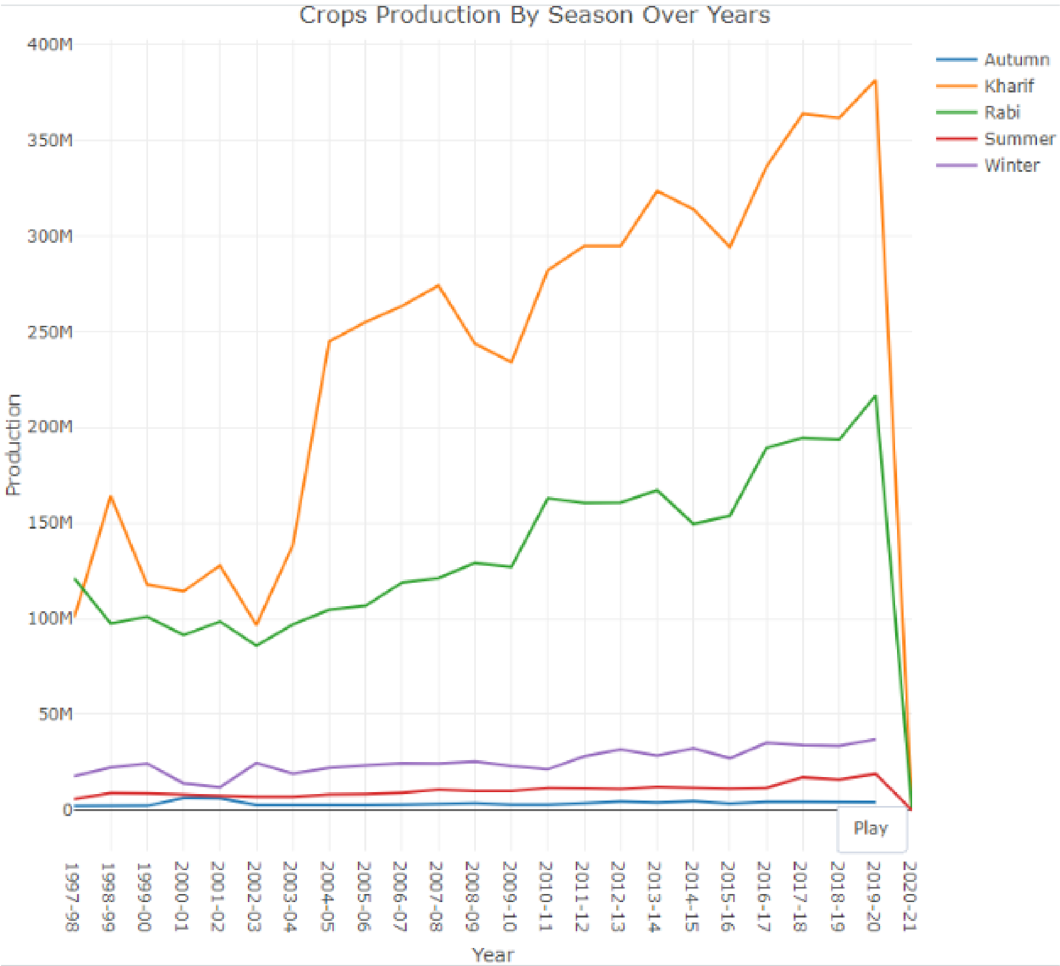


Fig 6.7.2 Over the years production

From the above line chart, we can infer that most of the production is happening in the season of kharif and rabi where the yield reached 380M in the year 2020-21 during kharif and 220M in rabi on the other hand autumn, summer and winter seasons production over the years has not reached 50M. So optimal crop recommended with the given parameters would increase the yield in these seasons is inferred.

CONCLUSION

In conclusion, the integration of data analytics into sustainable and efficient crop production practices is essential for addressing the challenges of climate change in agriculture.

Decision tree regression, Gradient Boosting, Random Forest, and Logistic Regression techniques are implemented on the input data to assess the best performance-yielding method.

These methods are compared using performance metrics. According to the analyses of cross-evaluation metrics, all four algorithms work well, but gradient boosting gives a slightly better accuracy score on test data than the other four models.

The proposed work can also be extended to analyze the climatic conditions and other factors for the crop and to increase crop production.

With the help of the high charter library visualizations on maps are made easier and we can see the insights the visualizations provide helps in making informed decisions over the coming years.

The use of data analytics also provides a valuable tool for researchers to understand the complex relationships between climate change and agricultural productivity, and to develop innovative solutions to address these challenges. By working together, farmers and researchers can ensure that sustainable and efficient crop production practices are implemented to mitigate the impact of climate change on food security, both now and in the future.

REFERENCES

[1] J. Majumdar, S. Naraseeyappa, and S. Ankalaki, "Analysis of agriculture data using data mining techniques: application of big data," *J Big Data*, vol. 4, no. 1, p. 20, Dec. 2017, doi: 10.1186/s40537-017-0077-4.

[2] N. Chergui and M. T. Kechadi, "Data analytics for crop management: a big data view," *J Big Data*, vol. 9, no. 1, p. 123, Dec. 2022, doi: 10.1186/s40537-022-00668-2.

[3] L. Kouadio and N. K. Newlands, "Building capacity for assessing spatial-based sustainability metrics in agriculture," *Decision Analytics*, vol. 2, no. 1, p. 2, Dec. 2015, doi: 10.1186/s40165-015-0011-9.

[4] S. Cai, D. Zuo, H. Wang, Z. Xu, G. Wang, and H. Yang, "Assessment of agricultural drought based on multi-source remote sensing data in a major grain producing area of Northwest China," *Agric Water Manag*, vol. 278, p. 108142, Mar. 2023, doi: 10.1016/j.agwat.2023.108142.

[5] A. I. Sourav and A. W. R. Emanuel, "Recent Trends of Big Data in Precision Agriculture: a Review," *IOP Conf Ser Mater Sci Eng*, vol. 1096, no. 1, p. 012081, Mar. 2021, doi: 10.1088/1757-899X/1096/1/012081.

[6] "How DBSCAN works and why should we use it? | by Kelvin Salton do Prado | Towards Data Science." https://towardsdatascience.com/how-dbscan-works-and-why-should-i-use-it-443b4a191c80 (accessed May 02, 2023).

[7] "python - How to find optimal parametrs for DBSCAN? - Stack Overflow." https://stackoverflow.com/questions/58983528/how-to-find-optimal-parametrs-for-dbscan (accessed May 02, 2023).

[8] "Highcharts for R users – Highcharts." https://www.highcharts.com/blog/tutorials/highcharts-for-r-users/ (accessed May 02, 2023).

[9] "An introduction to seaborn — seaborn 0.12.2 documentation." https://seaborn.pydata.org/tutorial/introduction.html (accessed May 02, 2023).

[10] "R Highcharter for Highcharts Data Visualization | DataCamp." https://www.datacamp.com/tutorial/data-visualization-highcharter-r (accessed May 02, 2023).

WORKLOG

Day	Date	Task Done
Day 1	15-02-2023	Collecting Research Paper on Project Topic
Day 2	16-02-2023	Collecting Research Paper on Project Topic
Day 3	17-02-2023	Collecting Research Paper on Project Topic
Day 4	20-02-2023	Collecting Research Paper on Project Topic
Day 5	21-02-2023	Deciding the Dataset
Day 6	22-02-2023	Deciding the Dataset
Day 7	23-02-2023	Deciding the Dataset
Day 8	24-02-2023	Finalizing the Dataset
Day 9	27-02-2023	Understanding the Dataset & EDA
Day 10	28-02-2023	Understanding the Dataset & EDA
Day 11	01-03-2023	Understanding the Dataset & EDA
Day 12	02-03-2023	Data Pre-processing
Day 13	03-03-2023	Data Pre-processing
Day 14	06-03-2023	Understanding seaborn visualization
Day 15	07-03-2023	Understanding seaborn visualization
Day 16	08-03-2023	Implementation summary statistics of data
Day 17	09-03-2023	Implementation summary statistics of data
Day 18	10-03-2023	Implementation summary statistics of data
Day 19	13-03-2023	Feature extraction
Day 20	14-03-2023	Feature extraction
Day 21	15-03-2023	Learning appropriate classification algorithms
Day 22	16-03-2023	Learning appropriate classification algorithms
Day 23	17-03-2023	Implementing Decision tree algorithm
Day 24	20-03-2023	Implementing Random Forest algorithm
Day 25	21-03-2023	Implementing Logistic Regression algorithm
Day 26	22-03-2023	Implementing Logistic Regression algorithm
Day 27	23-03-2023	Implementing Gradient boosting algorithm

Day 28	24-03-2023	Accuracy comparison
Day 29	27-03-2023	Visualization using seaborn
Day 30	28-03-2023	Visualization using seaborn
Day 31	29-03-2023	Plotting feature importance
Day 32	30-03-2023	Understanding DBscan Clustering
Day 33	31-03-2023	Understanding DBscan Clustering
Day 34	03-04-2023	Implementing DBscan and KMeans Clustering
Day 35	04-04-2023	Implementing DBscan and KMeans Clustering
Day 36	05-04-2023	Implementing DBscan and KMeans Clustering
Day 37	06-04-2023	Understanding Highcharter library
Day 38	10-04-2023	Understanding Highcharter library
Day 39	11-04-2023	Understanding Highcharter library
Day 40	12-04-2023	Understanding hcmap in R
Day 41	13-04-2023	Understanding hcmap in R
Day 42	17-04-2023	Visualizations using hcmap
Day 43	18-04-2023	Visualizations using hcmap
Day 44	19-04-2023	Visualizations using hcmap
Day 45	20-04-2023	Providing Insights on visualizations
Day 46	21-04-2023	Providing Insights on visualizations
Day 47	24-04-2023	Providing Insights on visualizations