A MINI PROJECT REPORT ON

"Crop Recommendation and growth analysis using Machine Learning Algorithms."

SUBMITTED TOWARDS THE FULFILLMENT OF THE REQUIREMENTS OF

BACHELOR OF ENGINEERING (B. Tech.)

Academic Year: 2023-24

By:

BTCOB81 TEJAS IPPAR

BTCOB88 KANAK JAISWAL

BTCOB91 SANJIT JHA

BTCOB92 PURNESH JOSHI

Under The Guidance of

Dr. K. Rajeswari

DEPARTMENT OF COMPUTER ENGINEERING,

PIMPRI CHINCHWAD COLLEGE OF ENGINEERING

SECTOR 26, NIGDI, PRADHIKARAN

PIMPRI CHINCHWAD COLLEGE OF ENGINEERING

DEPARTMENT OF COMPUTER ENGINEERING

CERTIFICATE

This is to certify that the project entitled

"Crop Recommendation and Growth Analysis using Machine Learning Algorithms."

is successfully carried out as a mini project successfully submitted by following students of "PCET's Pimpri Chinchwad College of Engineering, Nigdi, Pune-44**".**

Under the guidance of Prof. Dr. K. Rajeswari

In the fulfillment of the requirements for the B. Tech. (Computer Engineering)

BTCOB81 TEJAS IPPAR

BTCOB88 KANAK JAISWAL

BTCOB91 SANJIT JHA

BTCOB92 PURNESH JOSHI

(Prof. Dr. K. Rajeswari) (Prof. Dr. K. Rajeswari)

Guide Head,

Department of Computer Engineering

Index – I

Chapter no.	Sub-chapter no.		Title	Page no.
1			Introduction	1
	1.1		Problem Statement	1
	1.2		Project Objectives	1
	1.3		Motivation	1
	1.4		Literature Survey	1
2			Project Design	3
	2.1		Requirements and Resources	3
	2.2		Dataset Design	4
	2.3		Timeline Diagram	7
	2.4		Block Diagram	8
3			Results and Discussions	10
	3.1		Comparative analysis of all models	10
	3.2		GUI and Dashboard Screenshots	12
4			Conclusion	17
5			References	17

INDEX- II (FIGURES)

Figure no	Title	Page no.
1	Timeline Diagram	7
2	Block Diagram	8
3	Design Architecture	9
4	State wise Distribution	12
5	District wise Distribution	12
6	Crop wise Distribution	13
7	Time Series	13
8	Production	14
9	Crop Recommendation Form	14
10	Crop Prediction	15

INDEX – III (TABLES)

Table no	Title	Page no.
1	Corp dataset	6
2	Fert dataset	6
3	Merged Dataset	6
4	Comparative Analysis	10

Abstract

This project aims to empower farmers with data-driven insights for crop cultivation by implementing a crop recommendation system. Leveraging machine learning models, including the highly effective Random Forest, the system provides accurate and informed crop suggestions based on soil properties, environmental conditions, and economic viability. The interactive Power BI dashboard enhances user experience and decision-making, presenting data insights with precision. The comparative analysis of various models highlights Random Forest as the top performer, ensuring high accuracy, precision, and recall. The project's success lies in its potential to maximize crop yield, profitability, and sustainability while mitigating risks associated with farming. It represents a significant step towards addressing the challenges in agriculture and contributing to sustainable and profitable crop cultivation.

Chapter 1: Introduction

1.1 Problem Statement:

The problem statement involves the development of a system that recommends the most suitable crops for farmers to cultivate, considering various factors such as soil properties, environmental conditions, and economic viability. This system aims to empower farmers with data-driven insights, helping them make informed decisions before planting. The objective is to enhance crop yield, profitability, and sustainability while reducing risks associated with crop cultivation.

1.2 Project Objectives:

- Evaluate and recommend crops best suited to a particular farmer's land and conditions.
- Create an easy-to-use interface for farmers to input data and receive recommendations.
- Assist farmers in making informed decisions before cultivation to maximize crop yields and profitability.

1.3 Motivation:

Precision agriculture, a contemporary farming approach harnessing the power of soil characteristics, soil types, crop yield data, and real-time weather conditions, has emerged as a beacon of hope for the agricultural sector. Its core principle is simple yet transformative: recommending the most suitable crops for farmers based on intricate data analysis to maximize yield and profit while minimizing the risk of crop failures. In a world marked by the unpredictability of climate change and economic fluctuations, precision agriculture's recommendations play a pivotal role in empowering farmers to make informed decisions for their farming strategies. As the global agrarian crisis continues to pose challenges, the need for improved recommendation systems becomes ever more critical. These systems serve as a lifeline for farmers, offering guidance that can help mitigate risks, boost agricultural productivity, and ultimately contribute to sustainable and profitable crop cultivation.

1.4 Literature Survey:

[1] The agriculture industry's vital economic role and its importance as a primary profession for many populations are acknowledged. The paper focuses on leveraging machine learning algorithms to recommend crops for arid land, using features selected from areas with tropical climates where crops thrive. Five machine learning models were evaluated for crop recommendation, with "Random Forest" emerging as the most effective model, offering promising potential for improving agricultural productivity in arid regions.

[2] This paper introduces an ensemble learning-based crop planting recommendation algorithm, aiming to guide farmers in selecting the most suitable crops by considering environmental characteristics and soil element content. The study emphasizes the role of nitrogen (N), phosphorus (P), and potassium (K) elements in crop growth. It employs K-Nearest Neighbor (KNN), XGBoost, and Random Forest (RF) as weak learners, optimizing their parameters with Genetic Algorithm (GA). The ensemble learning model combines these weak learners using a soft voting mechanism. After evaluation on the Kaggle public dataset, the algorithm achieves an

impressive accuracy of 94.36%, demonstrating its potential for promoting scientific planting and increasing crop yields.

[3] In India, where a significant portion of the population relies on agriculture, crop production's quality is pivotal. The proposed IoT and ML system employs soil sensors to monitor and manage soil health, reducing the risk of soil degradation. It utilizes various sensors for soil temperature, moisture, pH, and NPK nutrient levels. Data collected by these sensors is analyzed using machine learning algorithms, particularly Random Forest, to recommend suitable crops. Additionally, the system incorporates a convolutional neural network for early disease risk detection in plants. This integrated approach offers a promising solution for improving crop production in India

[4] Agriculture is a vital component of the economies of developing countries like India, contributing significantly to GDP. The increasing population has raised food demand, but challenges in crop selection and input choices can lead to degradation in crop quality and profitability. Computational technologies have enabled the development of crop and fertilizer recommendation systems that consider factors like soil condition, water requirements, and market profitability. The literature presents "AgriRec," a machine learning-based recommendation algorithm that accurately suggests crops (95.85% accuracy) and fertilizers (92.11% accuracy) based on soil properties, water levels, and other factors. AgriRec outperforms existing benchmark approaches, offering a promising solution to enhance agricultural productivity.

Chapter 2: Project Design

2.1 Requirements and Resources

Hardware Requirements:

1. **Computer System:** We'll need a computer system capable of running Power BI Desktop and performing data analysis tasks. Ensure that it meets the minimum system requirements for Power BI.
2. **Storage:** We need sufficient storage space for data storage, data backups, and model outputs.

Software Requirements:

1. **Power BI:** Power BI Desktop is the primary software for creating interactive dashboards and visualizations. Ensure that we have the latest version installed.
2. **Data Analysis Tools:** Depending on our choice of machine learning models (Random Forest, Naive Bayes, Logistic Regression, IBK, J48.), we may need software packages or libraries for these models.
3. **Database Management:** If our data is stored in databases, we may need database management software to connect and retrieve data. Power BI allows connections to various data sources.
4. **Operating System:** Ensure that the operating system of our computer supports the required software and tools.

Resources:

1. **Data:** Our project's primary resource is the crop and fertilizers data, including information on humidity, temperature, pH, N, P, K required by a particular crop. Ensure that the data is clean, well-structured, and stored in a format compatible with Power BI.
2. **Models:** We'll need resources related to the machine learning models we plan to implement. This includes pre-trained models, if available, or the resources to train models using our dataset.
3. **Documentation:** Access to documentation for Power BI, machine learning libraries, and any other tools or software we plan to use. This will help in understanding and effectively utilizing these resources.

Requirements & Their Detailed Explanation:

- **Power BI Dashboard:**
  - Requirement: We need to create an interactive dashboard in Power BI.
  - Explanation: The dashboard will serve as the central interface for presenting data insights and machine learning model outputs. It should be user-friendly and provide a clear visualization of retail store data, including total revenue trends.
- **Data Analysis Tools:**
  - Requirement: Depending on the machine learning models chosen (Logistic Regression, Random Forest, Naïve Bayes, IBK, J48), we need the corresponding tools and libraries.

    - Explanation: Each machine learning model requires specific software and libraries for training, testing, and prediction. For example, we may need Python with scikit-learn for logistic regression and random forest, or Weka for J48.
- **Data Preprocessing:**
    - Requirement: Tools or scripts for data preprocessing and cleaning.
    - Explanation: Raw data often requires cleaning, transformation, and feature engineering to make it suitable for analysis. Data preprocessing tools and scripts are needed to ensure data quality.
- **Data Source Integration:**
    - Requirement: Capability to connect Power BI to the data source.
    - Explanation: Power BI should be able to retrieve data from the source database or files, and this capability should be set up properly.
- **Storage and Backup:**
    - Requirement: Adequate storage space and a backup strategy.
    - Explanation: We need sufficient storage to store the dataset, model outputs, and Power BI project files. Implement a backup strategy to avoid data loss.
- **Operating System Compatibility:**
    - Requirement: The chosen operating system must be compatible with Power BI and other required software.
    - Explanation: Ensure that the operating system can run Power BI and any additional software we plan to use without compatibility issues.

2.2 Dataset Design

For this project, we have used the Crop Recommendation Dataset available on kaggle. The dataset is made by augmenting and combining various publicly available datasets of India like weather, soil, etc. The dataset can be accessed here. This data is relatively simple with very few but useful features unlike the complicated features affecting the yield of the crop.

Data Fields:

1. %N (Nitrogen): Represents the nitrogen content in the soil, essential for plant growth, including structural development, metabolic processes, and chlorophyll synthesis. Inadequate nitrogen can lead to stunted growth and yellowing of plant tissues.
2. %P (Phosphorus): Captures soil phosphorus content, crucial for energy storage and transfer within plants, supporting growth and reproduction.
3. %K (Potassium): Denotes potassium content in the soil, fundamental for overall plant health and development.
4. Temperature: Measured in degrees Celsius, it influences photosynthesis and respiration processes in plants, impacting their growth.
5. Humidity: Relative humidity (expressed as a percentage) affects plant environments, with excessive humidity leading to mold, bacteria, and pest issues.
6. pH (pH Value): Measures soil acidity or alkalinity, influencing nutrient availability to plants. Extreme pH levels can hinder growth.
7. Rainfall: Represents the amount of rainfall in millimeters, vital for maintaining soil moisture levels crucial for healthy plant growth.
8. Label: The output variable with 22 distinct values, each corresponding to a different crop, providing insights into suitable crops for specific soil and environmental conditions.

The crop dataset offers valuable insights into essential factors such as humidity, temperature, and pH necessary for specific crops, while the fert dataset provides the corresponding and necessary N, P, and K values required by each crop. In the following discussion, we'll explore the process of merging these datasets to enhance the accuracy of our model:

Data Cleaning:

- The code defines a Python function called `change_case`, which is used to standardize the case and remove spaces from the crop and fertilizer names.
- It applies the `change_case` function to the `Crop` column in the `fert` DataFrame and the `label` column in the `crop` DataFrame, ensuring that the names are consistent in terms of case and spacing.

Data Standardization:

- The code standardizes certain crop names in the `fert` DataFrame to match the naming conventions in the `crop` DataFrame. For example, it replaces variations like `mungbeans` with `mungbean` and `lentils(masoordal)` with `lentil`.

Crop Names Extraction:

- The code extracts the unique crop names from both datasets, which are stored in the arrays `crop_names` and `crop_names_from_fert`.
- It identifies common crop names that exist in both datasets by comparing the two arrays.

Data Filtering:

- The script initializes two new DataFrames, `new_crop` and `new_fert`, which will be used to store filtered and standardized data.
- It iterates over the common crop names identified earlier and appends the corresponding data rows from the original `crop` and `fert` DataFrames to `new_crop` and `new_fert`, respectively.

Finally, the code exports the filtered and standardized crop data (in new_crop) to a new CSV file called 'MergeFileCrop.csv'

BEFORE:

	temperature	humidity	ph	rainfall	label
0	20.879744	82.002744	6.502985	202.935536	rice
1	21.770462	80.319644	7.038096	226.655537	rice
2	23.004459	82.320763	7.840207	263.964248	rice
3	26.491096	80.158363	6.980401	242.864034	rice
4	20.130175	81.604873	7.628473	262.717340	rice

Table 1. Corp.csv

	Unnamed: 0	Crop	N	P	K	pH
0	0	Rice	80	40	40	5.5
1	1	Jowar(Sorghum)	80	40	40	5.5
2	2	Barley(JAV)	70	40	45	5.5
3	3	Maize	80	40	20	5.5
4	4	Ragi(naachnnii)	50	40	20	5.5

Table 2. Fert.csv

AFTER:

	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

Table 3. MergeFileCorp.csv

2.3 Timeline Diagram

```
Project Start
│
├── Data Collection and Integration ( Crop details)
│   ├── Data Gathering
│   │
│   ├── Data Cleaning
│   │
│   ├── Data Transformation
│   │
├── Data Preprocessing
│   │
├── Model Selection
│   │
├── Model Development
│   ├── Random Forest
│   ├── Naive Bayes
│   ├── Logistic Regression
│   ├── IBK
│   ├── J48
│   │
├── Power BI Dashboard Design
│
├── Final Dashboard
│
├── Project Conclusion and Documentation
│
└── Project End
```

Fig. 1. Timeline Diagram

2.4 Block Diagram

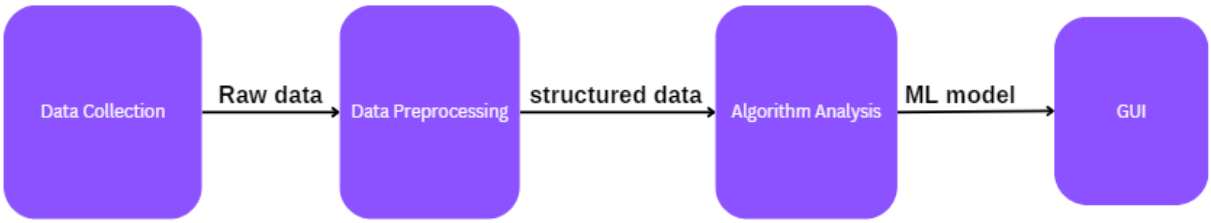


Fig. 2. Block Diagram

Explanation :

1. Data Collection:

Sources: Gather data from diverse agricultural sources, including crop yields, weather data, soil information, and farmer's cultivation practices. Two datasets were taken: crop and fert.

Methods: Employ data collection methods like remote sensing, IoT sensors, and farmer surveys to acquire real-time and historical data. APIs and connectors may be utilized for interfacing with relevant data sources.

2. Data Preprocessing:

Data Cleaning: Address data quality issues by handling missing or erroneous data, and eliminate duplications to ensure data accuracy.

Data Transformation: Normalize and scale data to ensure consistency, and apply feature engineering to extract valuable insights from the agricultural data..

Data Integration: Combine data from various sources, creating a unified dataset for comprehensive analysis.

3. Algorithm Analysis:

Machine Learning Models: Implement machine learning models such as regression for yield forecasting, clustering for soil type segmentation, and classification for crop disease prediction. These models provide insights and predictions for crop cultivation.

Statistical Analysis: Apply statistical analysis techniques, including descriptive statistics and hypothesis testing, to understand data distribution and validate assumptions in the context of agriculture.

Optimization Algorithms: Utilize optimization algorithms to enhance farming strategies, including crop selection, planting times, and resource allocation for improved efficiency and yield optimization.

4. Dashboard Visualization:

Visualization Tools: Utilize visualization tools, like custom-built dashboards or industry-specific solutions, to create visual representations of agricultural data.

Visualizations: Key visualizations include crop yield trends, soil condition analysis, disease prevalence maps, and seasonal weather patterns. These visualizations offer actionable insights to farmers and agricultural stakeholders.

Interactivity: Enhance the dashboard's interactivity by allowing users to explore data interactively, enabling them to make informed decisions and optimize crop cultivation practices based on data-driven insights.

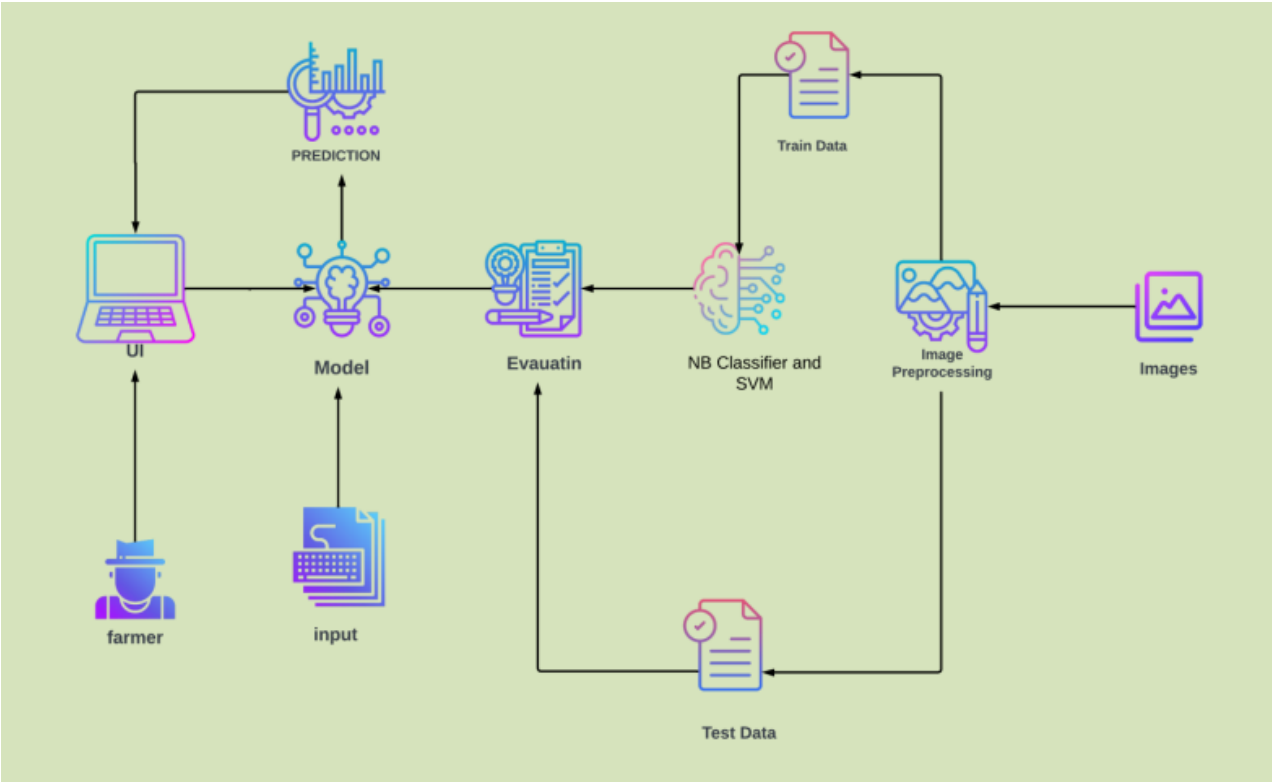


Fig. 3. Design Architecture

Chapter 3: Results & Discussion with comparative analysis

3.1 Comparative analysis of all models.

Models	Accuracy	Precision	Recall	F1-Score	MCC	ROC AUC	PRC AUC
Random Forest	99.36%	0.994	0.994	0.994	0.993	1	1
Naive Bayes	99.05%	0.99	0.99	0.99	0.99	1	0.99
Logistic Regression	97.91%	0.979	0.979	0.979	0.978	1	0.994
IBK	98.18%	0.982	0.982	0.982	0.981	0.99	0.966
J48	99.14%	0.991	0.991	0.991	0.991	0.997	0.986

Table 4. Comparative Analysis

Random Forest:

The Random Forest model exhibits exceptional performance with a high accuracy of 99.36%. This ensemble method combines multiple decision trees to enhance predictive accuracy and control overfitting. The low error metrics signify the model's effectiveness in minimizing both absolute and squared errors.

Naive Bayes:

Naive Bayes showcases excellent results with an accuracy of 99.05%. This probabilistic classifier assumes independence among features, and its success in this scenario suggests that feature independence might hold to a reasonable extent. The extremely low error metrics underscore the model's impressive performance.

Logistic Regression:

Logistic Regression demonstrates solid performance with an accuracy of 97.91%. This linear classifier models the probability of an instance belonging to a particular class. While slightly lower in accuracy compared to the previous models, the Iterative Classifier still performs well. The error metrics, although higher than other models, remain acceptable.

IBK (Instance-Based Classifier):

IBK, an instance-based classifier, achieves an accuracy of 98.18%. This model relies on the similarity between instances to make predictions. The Kappa statistic signifies a high level of agreement. Although the accuracy is slightly lower than some other models, it remains impressive. The error metrics are consistent with the model's overall accuracy.

J48 (Decision Tree):
J48, a decision tree-based classifier, achieves an accuracy of 99.14%. Decision trees are interpretable and can handle both numerical and categorical data effectively. The error metrics, while slightly higher than some other models, still reflect excellent performance.

Analysis :-

Accuracy: Random Forest achieved the highest accuracy at nearly 99.36%, followed by J48 and Naive Bayes. Logistic Regression has the lowest accuracy but is still at a respectable 97.91%.

Precision: Random Forest, Naive Bayes and J48 have nearly perfect precision scores, indicating that they make very few false-positive predictions. Logistic Regression and IBK also have good precision but are slightly lower than the top three.

Recall: Random Forest, Naive Bayes, J48 exhibit near-perfect recall, meaning they capture almost all actual positive instances. IBK and Logistic Regression have a slightly lower recall but are still strong.

F1-Score:Random Forest, Naive Bayes, J48 have the highest F1-Scores, indicating a balance between precision and recall. IBK and Logistic Regression also have good F1-Scores.

MCC (Matthews Correlation Coefficient)::Random Forest, Naive Bayes, J48 have very high MCC values, indicating strong overall performance.IBK and Logistic Regression, while still performing well, have lower MCC values.

ROC AUC (Receiver Operating Characteristic Area Under the Curve): Random Forest, Naive Bayes and Logistic Regression have perfect ROC AUC scores of 1.0, indicating excellent discriminative power. IBK and J48 have slightly lower ROC AUC scores but are still strong.

PRC AUC (Precision-Recall Curve Area Under the Curve): Random Forest again achieve perfect PRC AUC scores of 1.0, indicating excellent precision-recall trade-offs. Naive Bayes, Logistic Regression, IBK and J48 are slightly lower but still perform well.

In summary, Random Forest stands out as top performers in this classification task. It achieves near-perfect accuracy, precision, recall, F1-Score, MCC, and area under the ROC and PRC curves.

3.2 Screenshots including GUI and Dashboard : Explain Dashboard components

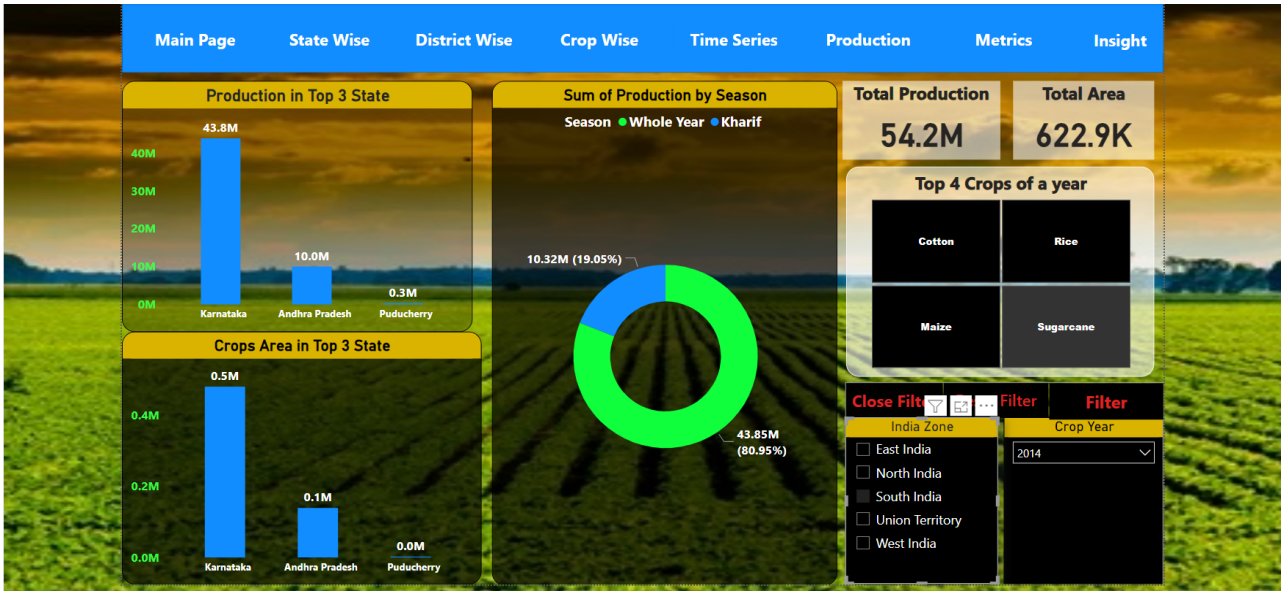


Fig. 4. State wise distribution

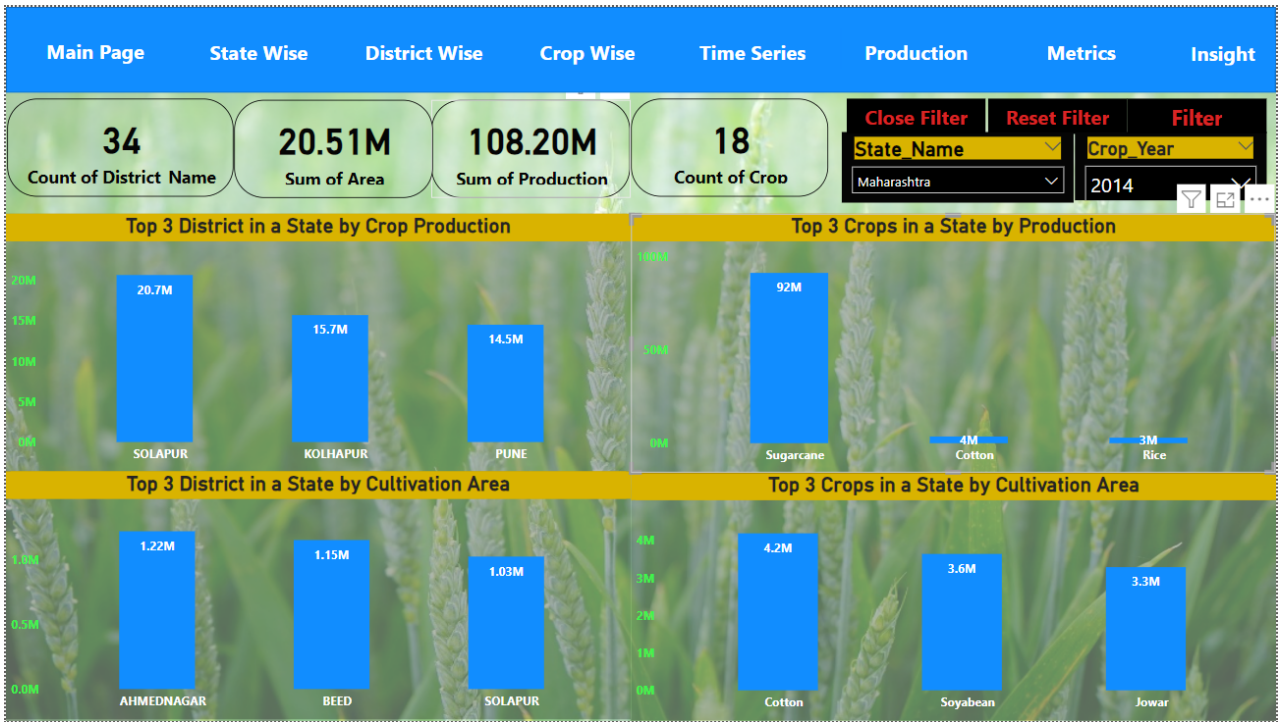


Fig. 5. District wise distribution

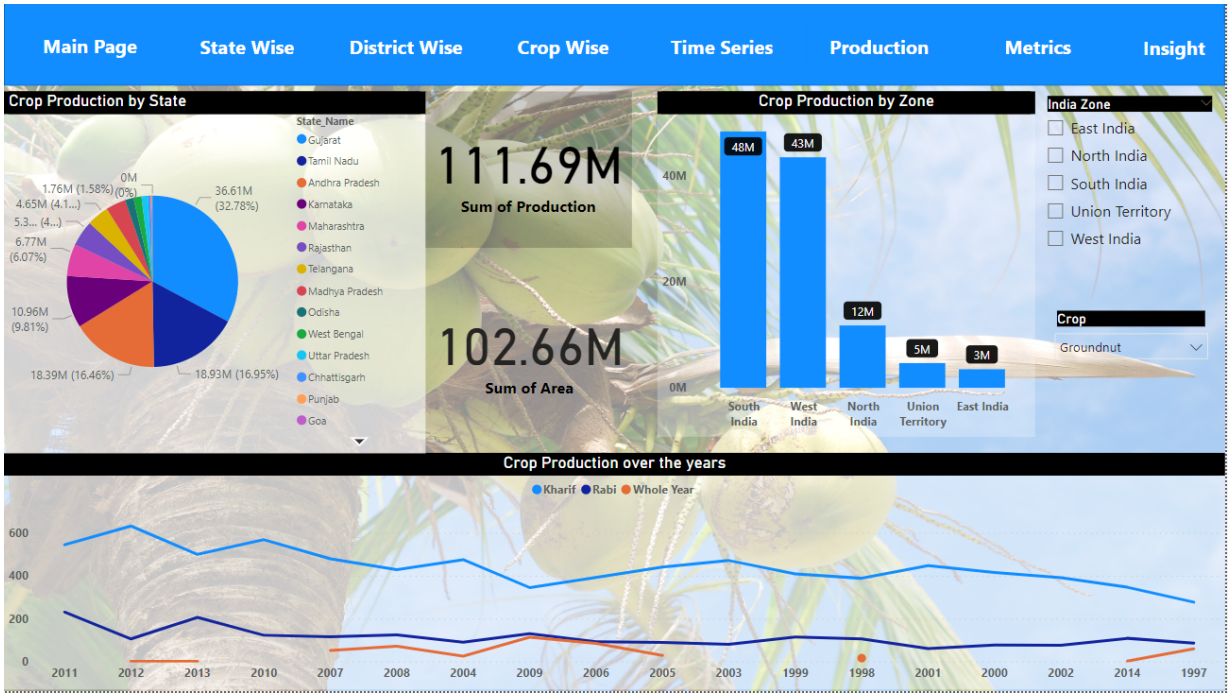
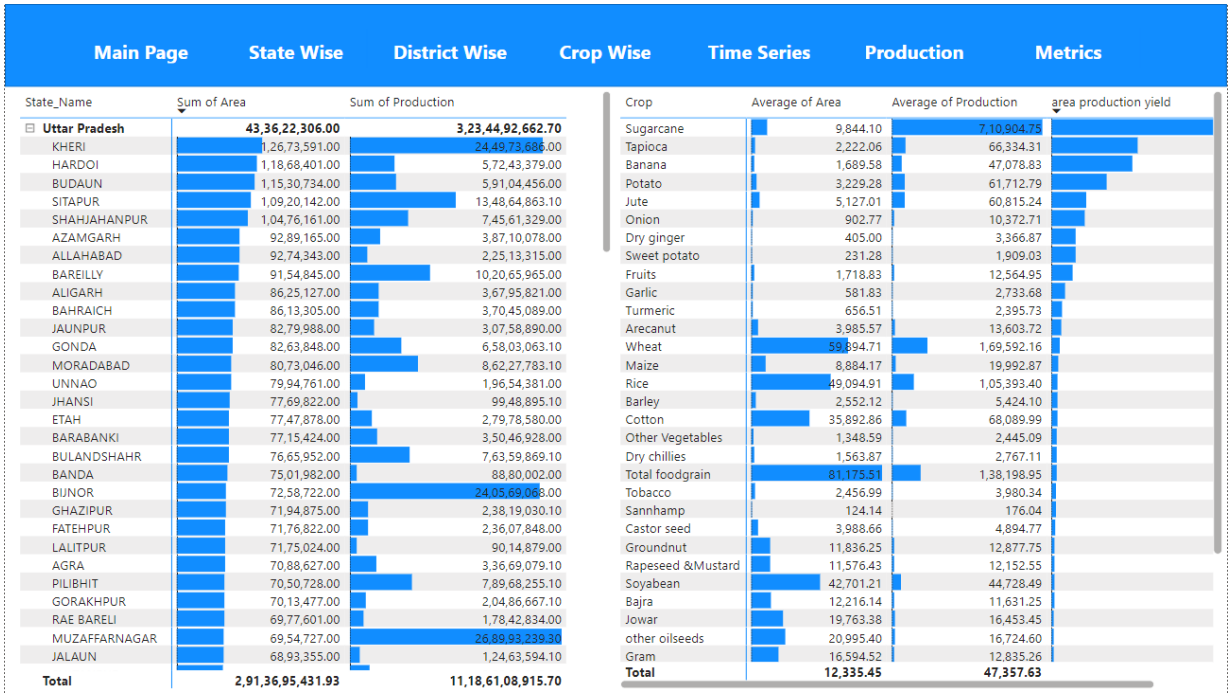


Fig. 6. Crop wise Distribution

Fig. 7. Time series



State_Name	Sum of Area	Sum of Production
Uttar Pradesh	**43,36,22,306.00**	**3,23,44,92,662.70**
KHERI	1,26,73,591.00	24,49,73,686.00
HARDOI	1,18,68,401.00	5,72,43,379.00
BUDAUN	1,15,30,734.00	5,91,04,456.00
SITAPUR	1,09,20,142.00	13,48,64,863.10
SHAHJAHANPUR	1,04,76,161.00	7,45,61,329.00
AZAMGARH	92,89,165.00	3,87,10,078.00
ALLAHABAD	92,74,343.00	2,25,13,315.00
BAREILLY	91,54,845.00	10,20,65,965.00
ALIGARH	86,25,127.00	3,67,95,821.00
BAHRAICH	86,13,305.00	3,70,45,089.00
JAUNPUR	82,79,988.00	3,07,58,890.00
GONDA	82,63,848.00	6,58,03,063.10
MORADABAD	80,73,046.00	8,62,27,783.10
UNNAO	79,94,761.00	1,96,54,381.00
JHANSI	77,69,822.00	99,48,895.10
ETAH	77,47,878.00	2,79,78,580.00
BARABANKI	77,15,424.00	3,50,46,928.00
BULANDSHAHR	76,65,952.00	7,63,59,869.10
BANDA	75,01,982.00	88,80,002.00
BIJNOR	72,58,722.00	24,05,69,068.00
GHAZIPUR	71,94,875.00	2,38,19,030.10
FATEHPUR	71,76,822.00	2,36,07,848.00
LALITPUR	71,75,024.00	90,14,879.00
AGRA	70,88,627.00	3,36,69,079.10
PILIBHIT	70,50,728.00	7,89,68,255.10
GORAKHPUR	70,13,477.00	2,04,86,667.10
RAE BARELI	69,77,601.00	1,78,42,834.00
MUZAFFARNAGAR	69,54,727.00	26,89,93,239.30
JALAUN	68,93,355.00	1,24,63,594.10
Total	**2,91,36,95,431.93**	**11,18,61,08,915.70**

Crop	Average of Area	Average of Production	area production yield
Sugarcane	9,844.10	7,10,904.75	
Tapioca	2,222.06	66,334.31	
Banana	1,689.58	47,078.83	
Potato	3,229.28	61,712.79	
Jute	5,127.01	60,815.24	
Onion	902.77	10,372.71	
Dry ginger	405.00	3,366.87	
Sweet potato	231.28	1,909.03	
Fruits	1,718.83	12,564.95	
Garlic	581.83	2,733.68	
Turmeric	656.51	2,395.73	
Arecanut	3,985.57	13,603.72	
Wheat	59,394.71	1,69,592.16	
Maize	8,884.17	19,992.87	
Rice	49,094.91	1,05,393.40	
Barley	2,552.12	5,424.10	
Cotton	35,892.86	68,089.99	
Other Vegetables	1,348.59	2,445.09	
Dry chillies	1,563.87	2,767.11	
Total foodgrain	81,175.51	1,38,198.95	
Tobacco	2,456.99	3,980.34	
Sannhamp	124.14	176.04	
Castor seed	3,988.66	4,894.77	
Groundnut	11,836.25	12,877.75	
Rapeseed &Mustard	11,576.43	12,152.55	
Soyabean	42,701.21	44,728.49	
Bajra	12,216.14	11,631.25	
Jowar	19,763.38	16,453.45	
other oilseeds	20,995.40	16,724.60	
Gram	16,594.52	12,835.26	
Total	**12,335.45**	**47,357.63**	

Fig. 8. Production

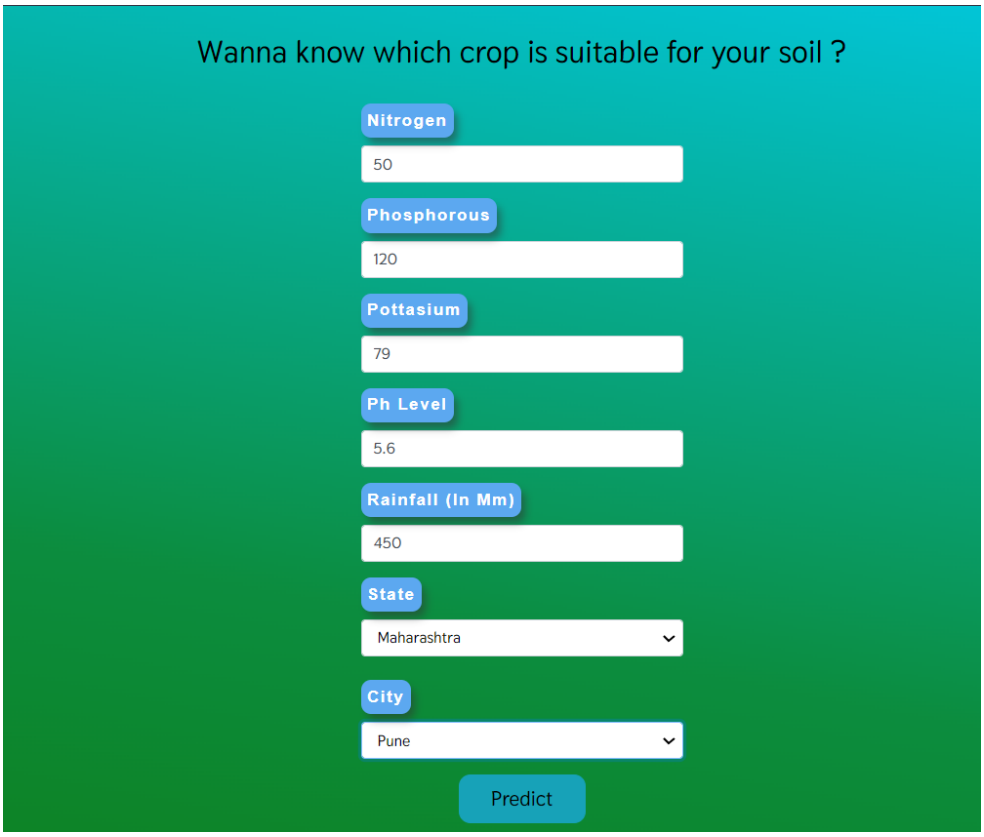


Fig. 9. Crop Recommendation Form

Fig. 10. Crop Prediction

Explain Dashboard components –
Power BI provides a wide range of dashboard components to help you visualize and analyze your data effectively. Here are some of the key dashboard components of Power BI:

1. Visualizations:
Power BI offers various types of visualizations, including bar charts, column charts, pie charts, line charts, scatter plots, maps, tables, and more. These visualizations help you represent data in different ways.

2. Cards:
Card visuals display a single, prominent metric or KPI (Key Performance Indicator). You can use them to highlight important numbers such as total sales, revenue, or profit**.**

3. Slicers:
Slicers allow users to filter data dynamically. They provide a way to interact with your data by selecting specific categories, date ranges, or other filter criteria.

4. Tables and Matrices:
Tables and matrices present tabular data. You can add columns and rows to display detailed information and include conditional formatting to emphasize certain values.

5. Gauges:
Gauges display single values within a range, making it easy to monitor performance against predefined thresholds. Gauges are useful for displaying progress or percentages.

6. Images and Shapes:
You can add images, icons, and shapes to your dashboard for branding, decoration, or to add context to your data.

7. KPI Indicators:
KPI indicators visually represent key performance metrics. They often use icons or traffic light symbols to indicate performance levels.

8. Text Boxes:
Text boxes allow you to add titles, subtitles, and explanations to your dashboard to provide context and guidance to users.

9. Buttons:
Buttons provide interactivity by allowing users to trigger actions or navigate between different report pages or dashboards.

10. Custom Visuals:
Power BI supports custom visuals created by the community or developed in-house. These can extend the range of visualizations available in Power BI.

Chapter 4: Conclusion

In conclusion, the Random Forest model has demonstrated outstanding performance in various key metrics, including accuracy, precision, recall, Matthews Correlation Coefficient (MCC), and Area Under the Curve (AUC). Its versatility and adaptability make it the optimal choice for our crop recommendation system, significantly enhancing the reliability of suggestions provided to farmers. By leveraging the Random Forest model, we are not only contributing to improved crop yields but also empowering farmers with data-driven insights, thereby promoting informed decision-making in agriculture. This robust model stands as a pivotal asset in advancing agricultural practices and fostering sustainable growth in the farming sector.

References

[1]Suresh, Sankaranarayanan. (2023). Machine Learning Enabled Crop Recommendation System for Arid Land.

[2]Guangyang, Deng., Chen, Dong., Jiabo, Chen., Jiaxing, Gao., Chunqiang, Li. (2023). Crop planting recommendation algorithm based on ensemble learning.

[3]P Kumar., Kusum, Lata., Sushant, Jhingran. (2023). Crop recommendation using machine learning algorithms. Available from: 10.1109/ICDT57929.2023.10151325.

[4]Krupa, Patel. (2023). Multi-criteria Agriculture Recommendation System using Machine Learning for Crop and Fertilizers Prediction.