<u>Second year Mini Project Report</u>

Submitted in partial fulfillment of the requirements of the degree

BACHELOR OF ENGINEERING IN COMPUTER ENGINEERING

By

Shreerang Vaidya D7B/59

Samarth Nilkanth D7B/38

Jatin Navani D7B/34

Anjali Thakrani D7B/57

Supervisor

Prof. Mrs. Yugchhaya Galphat

Since 1962

Department of Computer Engineering

Vivekanand Education Society's Institute of Technology

HAMC, Collector's Colony, Chembur,

Mumbai-400074

University of Mumbai

(AY 2023-24)

CERTIFICATE

This is to certify that the Mini Project entitled **“KisanSevak: Data-Driven Agriculture for Optimal Yields and Product Procurement”** is a bonafide work of **Shreerang Vaidya (59), Samarth Nilkanth (38), Jatin Navani (34), Anjali Thakrani (57)** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **“Bachelor of Engineering”** in **“Computer Engineering” .**

(Prof.__________________)
Supervisor

(Prof.__________________)
Head of Department

(Prof.__________________)
Principal

Mini Project Approval

This Mini Project entitled "KisanSevak: Data-Driven Agriculture for Optimal Yields and Product Procurement" by **Shreerang Vaidya (59), Samarth Nilkanth (38), Jatin Navani (34), Anjali Thakrani (57)** is approved for the degree of **Bachelor of Engineering** in **Computer Engineering.**

Examiners

1………………………………………
(Internal Examiner Name & Sign)

2………………………………………
(External Examiner name & Sign)

Date:

Place:

Contents

Abstract

"Kisan Sevak" is an innovative initiative aimed at enhancing the livelihood of farmers through targeted interventions. The primary focus is on leveraging the power of Machine Learning to optimise agricultural output. The initiative also seeks to establish convenient channels for farmers to procure essential resources and effectively market their good quality crops at competitive prices. The core objective of the "Kisan Sevak" app is to address the multifaceted challenges encountered by farmers throughout the entire agricultural lifecycle, from initial setup i.e from buying of seed crops to final production of grains.This interactive mobile and web-based application serves as a comprehensive guide, assisting farmers across various stages of the farming process. "Kisan Sevak" harnesses the capabilities of curated databases and advanced algorithms to deliver accurate and tailored information to farmers. The overarching aim is to maximise the yield potential of their farmland, translating into higher profitability and increased quality of the grains.

Acknowledgments

We would like to thank and express gratitude to all those who contributed and supported us to plan our project smoothly and successfully.

We would like to express our gratitude towards Dr. J. M. Nair, Principal of V.E.S.I.T for her immense support and motivation.

Firstly, we would like to thank Dr. Nupur Giri, Head of Department, Computer Engineering of V.E.S. Institute of Technology, for her guidance. We are whole-heartedly thankful to her for giving us their valuable time and knowledge to make us understand the executing process and hence providing a systemic planning of our project in time.

We would like to thank our project coordinator Mrs. Vidya Zope under whose guidance, we could learn many things. Not just this, she motivated us and strengthened our confidence in the entire execution process.

We express our immense gratitude to Mrs. Yugchhaya Galphat for her constant guidance and valuable suggestions which made us complete the execution of our project successfully. Her guidance and pattern of teaching made us capable enough to plan the project systematically and efficiently.

We would also like to extend our gratitude to all the faculty members who have not just been a constant source of support, but also encouraged us for timely completion of assigned execution activity.

Lastly, we would like to acknowledge our classmates, who have also provided us with every possible support and learning to execute our project efficiently.

List of Abbreviations

Sr No.	Abbreviation	Full Form
1	LSTM	Long Short-Term Memory
2	SVM	Support Vector Machine
3	RMSE	Root Mean Square Error
4	MAE	Mean Absolute Error
5	KNN	K-Nearest Neighbours
6	LDA	Linear Discriminant Analysis
7	LR	Logistic Regression

List of Figures

List of Tables

1. Introduction

1.1. Introduction

India has long been recognized as a global agricultural powerhouse, producing substantial quantities of wheat, rice, cotton, milk, pulses, and spices. Its agricultural sector is not only a vital contributor to the national economy but also a cornerstone of the livelihoods of millions of farmers. However, despite this agricultural abundance, Indian farmers confront a myriad of challenges that impede their socio-economic well-being and threaten their very livelihoods.

One of the biggest challenges faced by Indian farmers is small and fragmented landholdings. The average landholding size in India is just 1.1 hectares, which is significantly lower than the global average of 4.4 hectares. This fragmentation of landholdings makes it difficult for farmers to achieve economies of scale and adopt modern farming practices. Another major challenge is dependence on monsoon rainfall. Agriculture in India is heavily dependent on monsoon rainfall, which is often erratic and unpredictable. This uncertainty in rainfall patterns can lead to crop failure and financial losses for farmers.

Indian farmers also face a number of other challenges, including lack of access to credit, high input costs, and low output prices. These challenges have a significant impact on the lives of farmers, forcing many to live in poverty and debt.

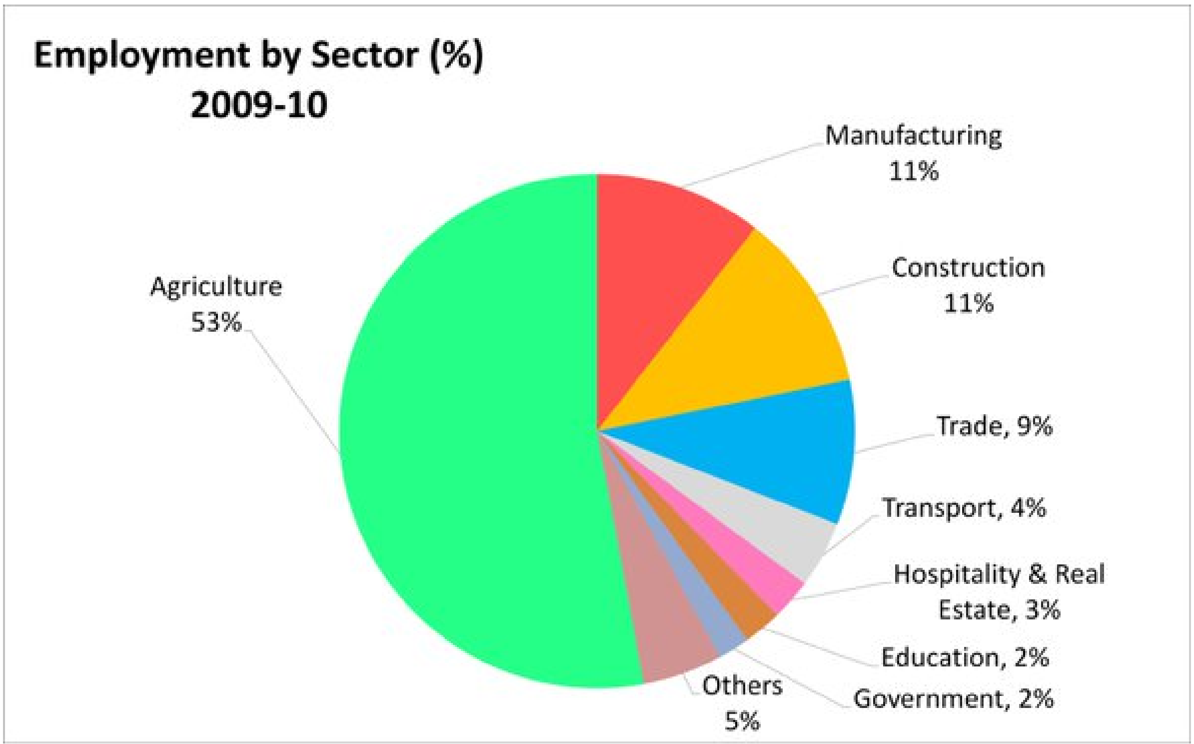


Figure 1.1. Employment % in India by Sector

1.2. Motivation

The motivation behind this project stems from the urgent need to alleviate the struggles faced by Indian farmers. The challenges they face include socio-economic insecurity and chronic indebtedness. These issues are often exacerbated by a variety of factors, such as middlemen who insert themselves between farmers and the market, inflating prices and reducing farmers' profits. Additionally, the high costs of agricultural inputs, including seeds and equipment, can place a heavy financial burden on already financially vulnerable farming communities. Moreover, the absence of exposure to modern farming technologies and best practices further hampers their ability to maximise crop production and access lucrative markets.

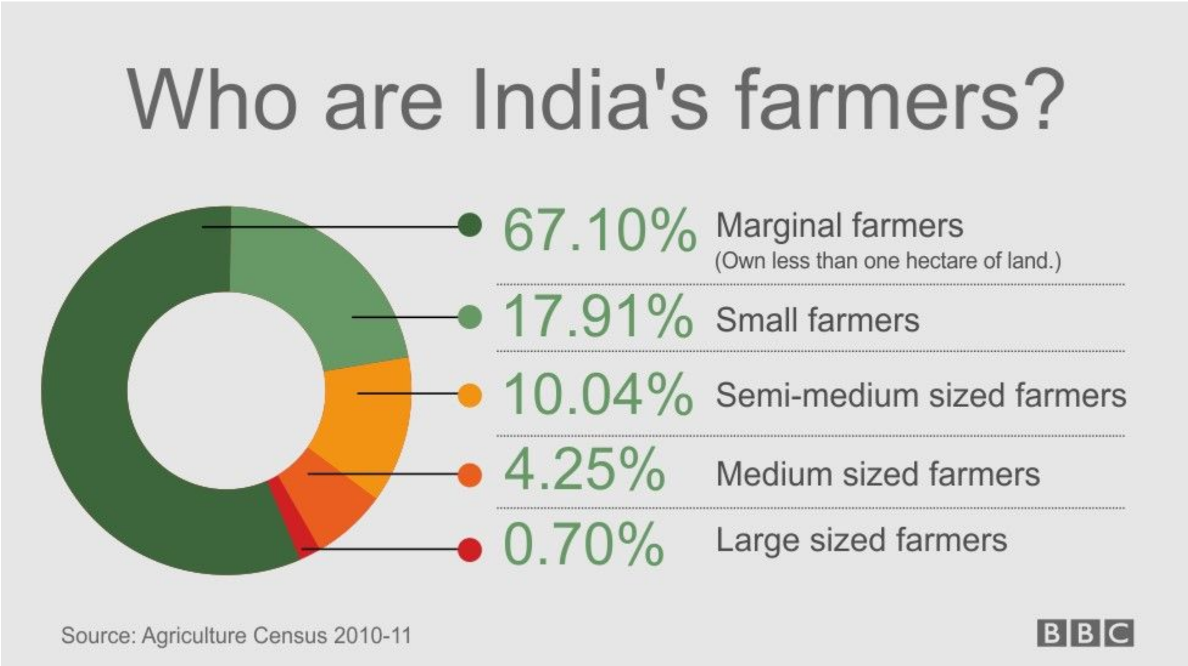


Figure 1.2. Farmer % by farm size

1.3. Problem Statement & Objectives

Despite India's status as a leading global agricultural producer, the nation's farmers encounter a multitude of challenges that undermine their livelihoods and the sustainability of the agricultural sector. These challenges include inadequate market linkages that force farmers to rely on middlemen, leading to reduced profits and inflated consumer prices. The lack of guidance in crop selection based on local conditions results in suboptimal yields and financial instability.

Furthermore, unpredictable market price fluctuations make it challenging for farmers to set fair prices for their produce, contributing to income volatility. Insufficient access to modern farming practices, sustainable techniques, and vital policy information hinders productivity and growth. Additionally, a significant proportion of farmers grapple with chronic indebtedness due to factors like high input costs and limited credit availability. Addressing these challenges is crucial to empower Indian farmers, enhance their economic well-being, and ensure the long-term sustainability of the country's vital agricultural sector.

Our current objective is to provide farmers with an easy-to-use ML tool to help them make better decisions about their crops using various models, and to eventually also provide them with a market to buy better fertilisers, tools, etc.

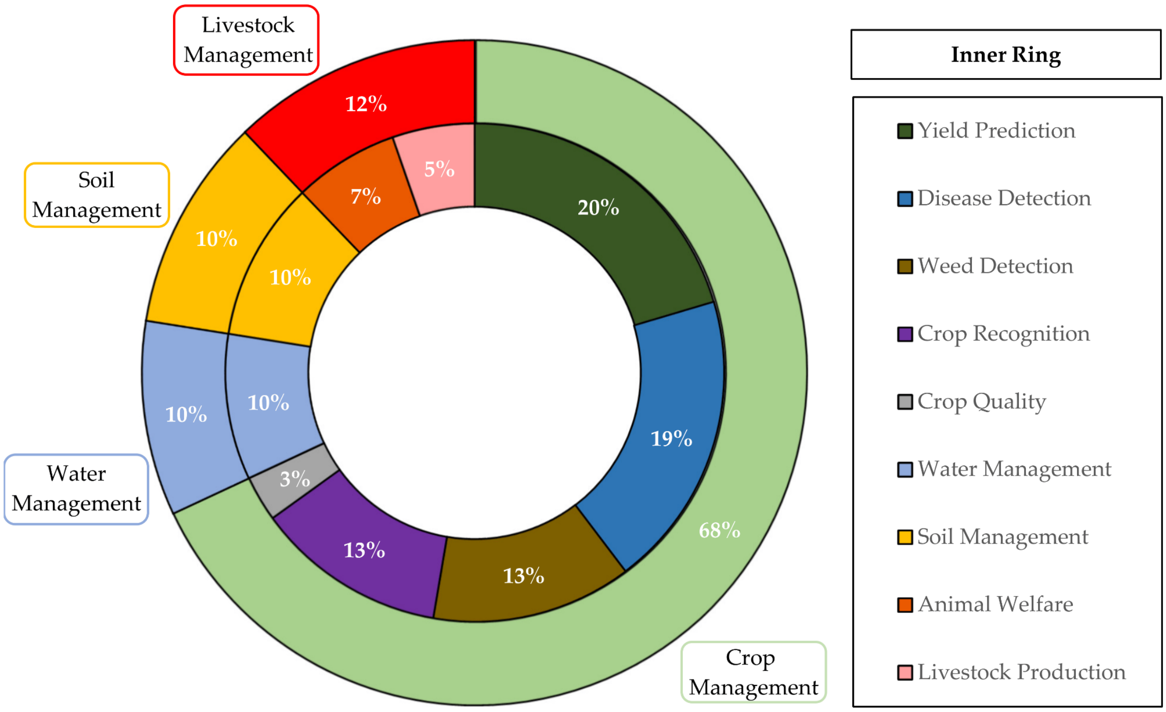


Figure 1.3. Various applications of ML in agriculture

1.4. Organization of the Report

In this report, we further discuss the following points:

- Literature survey of existing systems
- Limitations of existing systems
- Mini project contribution
- The proposed system
- Details of hardware and software used
- Progress
- Conclusion

2. Literature Survey

2.1. Survey of Existing System

1. **Crop Recommendation System Using ML**

Abstract:
This paper proposes a crop recommendation system using ML algorithms. The system considers various factors such as soil properties, climatic conditions, and market prices to recommend the most suitable crops for a given region. The system is implemented using a variety of ML algorithms, including SVM, random forest, and decision tree. The system is evaluated on a real-world dataset of agricultural data from India. The results show that the system can achieve an accuracy of over 90% in recommending the most suitable crops for a given region.

Inference:
The results of this study suggest that machine learning algorithms can be used to accurately predict soil fertility and crop yield. The authors found that the Random Forest performed the best overall.

2. **AI-based Market Intelligence Systems for Farmer Collectives: A Case-study from India**

Abstract:
This paper presents a case study of an AI-based market intelligence system that was developed to help a network of farmer cooperatives in India growing soybeans. The system uses historical market data to forecast soybean prices and provides farmers with recommendations on when to sell their produce.

Inference:
The paper advocates for the utilisation of LSTM (Long Short-Term Memory) neural networks to capture complex market dynamics and trends, benefiting farmer collectives with insightful pricing information.

3. **Evaluating Machine Learning Algorithms for Predicting Maize Yield under Conservation Agriculture in Eastern and Southern Africa**

Abstract:
This study evaluated the performance of various machine learning algorithms for predicting maize yield under conservation agriculture in Eastern and Southern Africa. The authors used a dataset of maize yield data from 120 sites in the region, collected over a period of three years. The dataset included a variety of factors, such as soil properties, climatic conditions, and farming practices.

Inference:
The results of this study suggest that machine learning algorithms can be used to accurately predict maize yield. The authors found that the LDA and LR algorithms performed the best, with accuracies of over 90%.

4. **Comparative Analysis of Soil Properties to Predict Fertility and Crop Yield using Machine Learning Algorithms**

Abstract:
The authors used a dataset of 120 soil samples collected from different regions of India. The soil samples were analysed for various properties, such as pH, nitrogen content, phosphorus content, and potassium content. The crop yield data was collected from farmers in the same regions.

Inference:
The authors found that the Random Forest performed the best for yield prediction, with an accuracy of 95.83%.

Together, these papers exemplify the significance of machine learning and artificial intelligence in modern agriculture, addressing various aspects from crop recommendations and soil health evaluation to market price prediction and yield estimation. This research is crucial in revolutionising farming practices and ensuring sustainable and profitable agricultural outcomes.

Existing systems

- **Government of India - Farmers' Portal**: https://farmer.gov.in/
  This is a government project for educating farmers in terms of latest data and updated information regarding various fields.

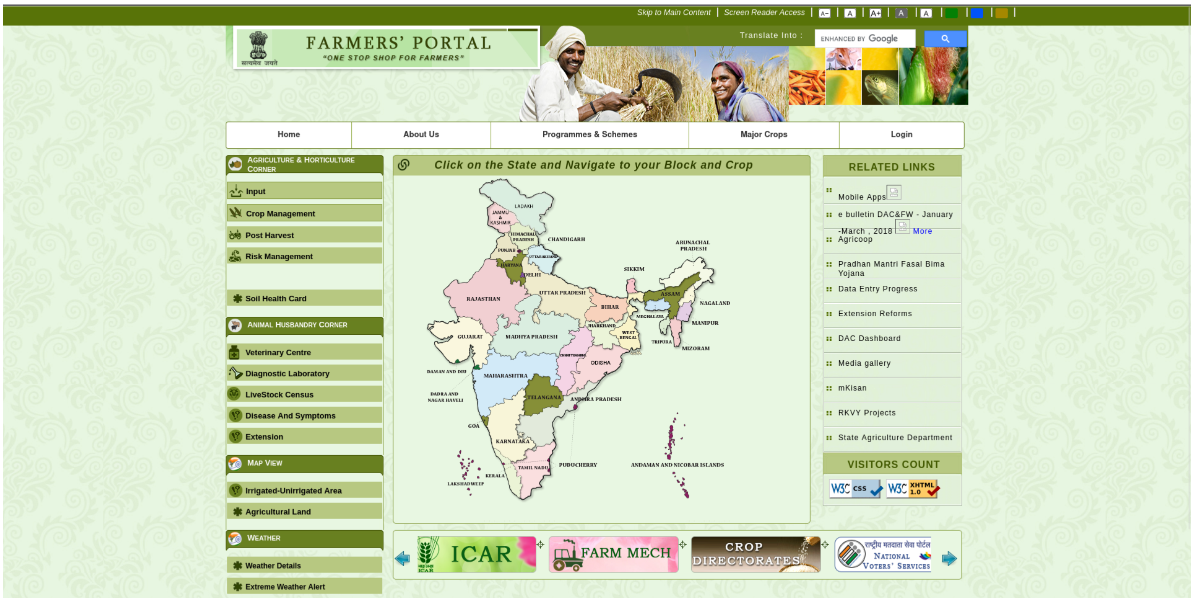


Figure 2.1. Existing System: Farmers' Portal

- **FarmOS**: https://farmos.org/ - This is a free and open-source web-based platform that uses AI to help farmers with a variety of tasks, such as crop monitoring, pest and disease detection, and yield prediction. The platform has been used by over 10,000 farmers in the United States, and it has helped farmers to reduce their costs and improve their yields.

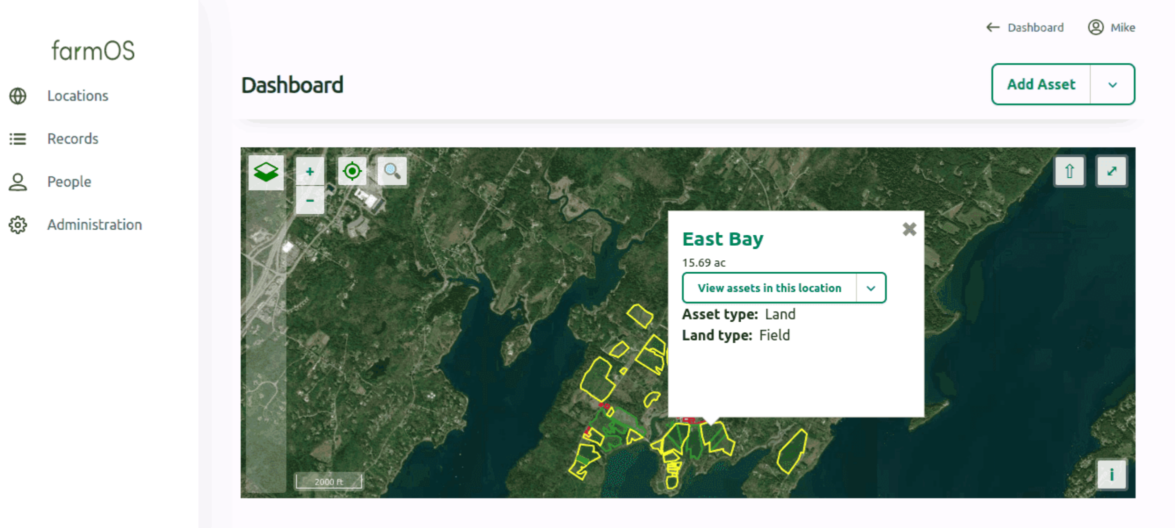


Figure 2.2. Existing System: FarmOS

2.2. Limitation Existing system or research gap

The absence of a comprehensive combinational system for crop suggestions, price predictions, and recommended agricultural measures presents a pressing issue in the farming industry. Farmers are vulnerable to significant financial losses due to unpredictable fluctuations in market prices. Furthermore, the lack of a user-friendly user interface exacerbates the problem, especially for farmers with limited education.

To address this, it is imperative to develop a holistic solution that integrates data-driven insights, market trends, and personalised recommendations. Such a system could help farmers make informed decisions about what crops to plant and when, predict market price trends, and provide step-by-step guidance on best practices for their crops. It must prioritise simplicity and accessibility to cater to the needs of uneducated farmers, ensuring that vital information is easily understood and actionable. By bridging these gaps, we can empower farmers, mitigate financial risks, and enhance the overall efficiency and sustainability of the agricultural sector.

Furthermore, there are no indigenous solutions that provide an accuracy and service comparable to local solutions from developed countries. This is due to a number of factors, including the lack of investment in research and development, the shortage of skilled personnel, and the limited access to data.

As a result, farmers in developing countries are often forced to rely on outdated and inaccurate information, which puts them at a significant disadvantage. A holistic solution that integrates data-driven insights, market trends, and personalised recommendations could help to bridge this gap and level the playing field for farmers in all countries.

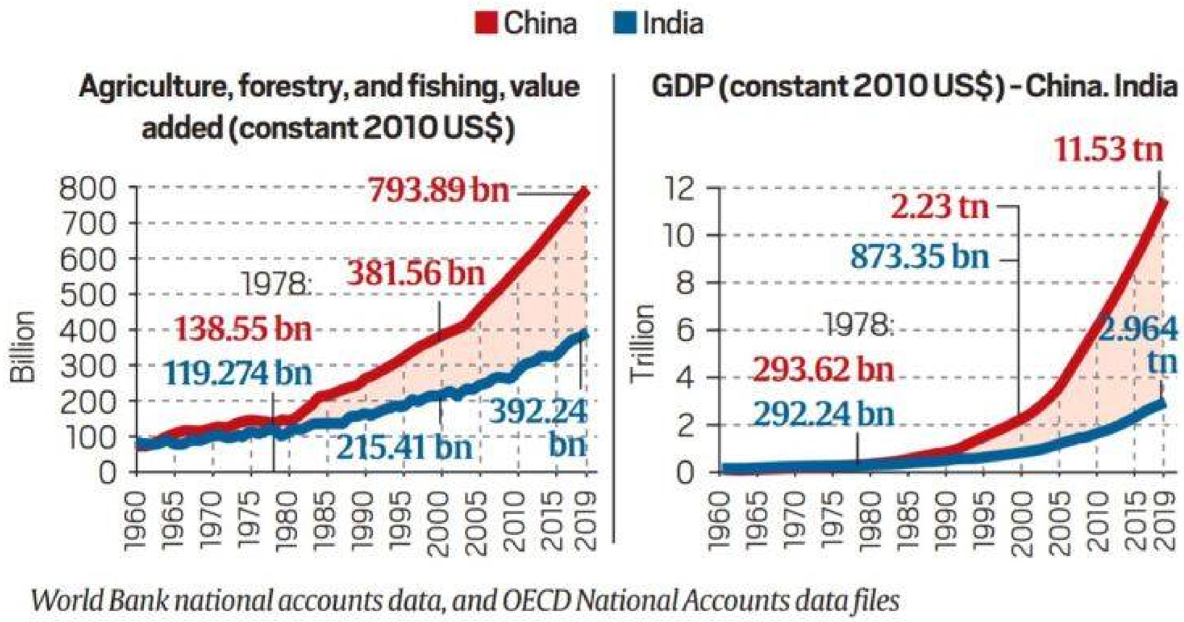


Figure 2.3. Contribution of Agriculture and GDP Comparison - India and China

2.3. Mini Project Contribution

A project that helps hard-working farmers to leverage the power of machine learning in order to have better yields and plant the right crops at the right time, not to mention, let the farmers know the right time to sell crops by predicting market prices for various crops would have a number of significant contributions to STEM and society.

Contribution to STEM

- The project would help to advance the field of agricultural machine learning, which is a rapidly growing field with the potential to revolutionise the way that food is produced.
- The project would provide valuable data and insights that could be used to improve the accuracy and performance of agricultural machine learning models.
- The project would train and develop a new generation of STEM professionals who are skilled in applying machine learning to agricultural problems.

Contribution to society

- The project would help to increase agricultural productivity and yields, which could help to reduce food insecurity and hunger around the world.
- The project would help farmers to make more informed decisions about what to plant and when, which could help to improve their profits and livelihoods.
- The project would help farmers to reduce their environmental impact, by helping them to use resources more efficiently and reduce waste.

Overall, a project that helps hard-working farmers to leverage the power of machine learning would have a number of positive and significant contributions to STEM and society.

3. Proposed System

3.1. Introduction

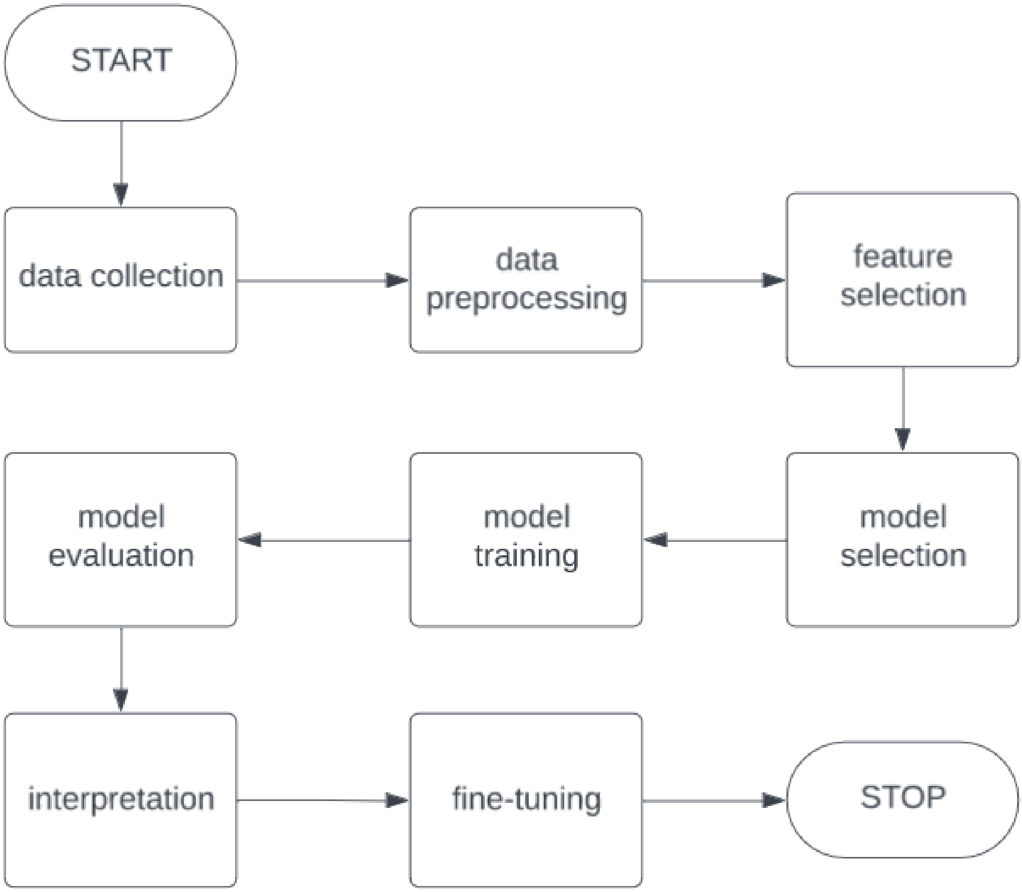
The idea is to create an interface for farmers in order to help them plan their cultivation strategy and sales. The user (farmer) can make an account and place required sensors in their field. After the initial steps no work is required from the user.

The app will recommend solutions to the user as per real-time predictions based on sensor data (currently simulated) and regional factors. Prediction graphs regarding fertility of the field, market prices, etc. can be obtained, as well as recommendations for future cultivation. If the user clicks on a graph, more details can be generated and displayed with more parameters.

The user interface (UI) of KisanSevak is designed with simplicity in mind. It is extremely user-friendly and lightweight, ensuring that farmers, regardless of their tech-savviness, can navigate the app effortlessly. The intuitive design allows users to access critical information and features without any hassle.

In addition to the predictive and analytical features, KisanSevak offers a marketplace module. Farmers can conveniently search for and purchase fertilisers, seeds, and other agricultural essentials from credible sources. The module's extensive database provides detailed information about these products. Users can click on any item to access more information and make purchases directly through the app. All transaction records are securely saved to the user's account, offering a comprehensive overview of their agricultural purchases and expenses.

3.2. Architecture/Framework



3.1.a Flowchart of project architecture - ML

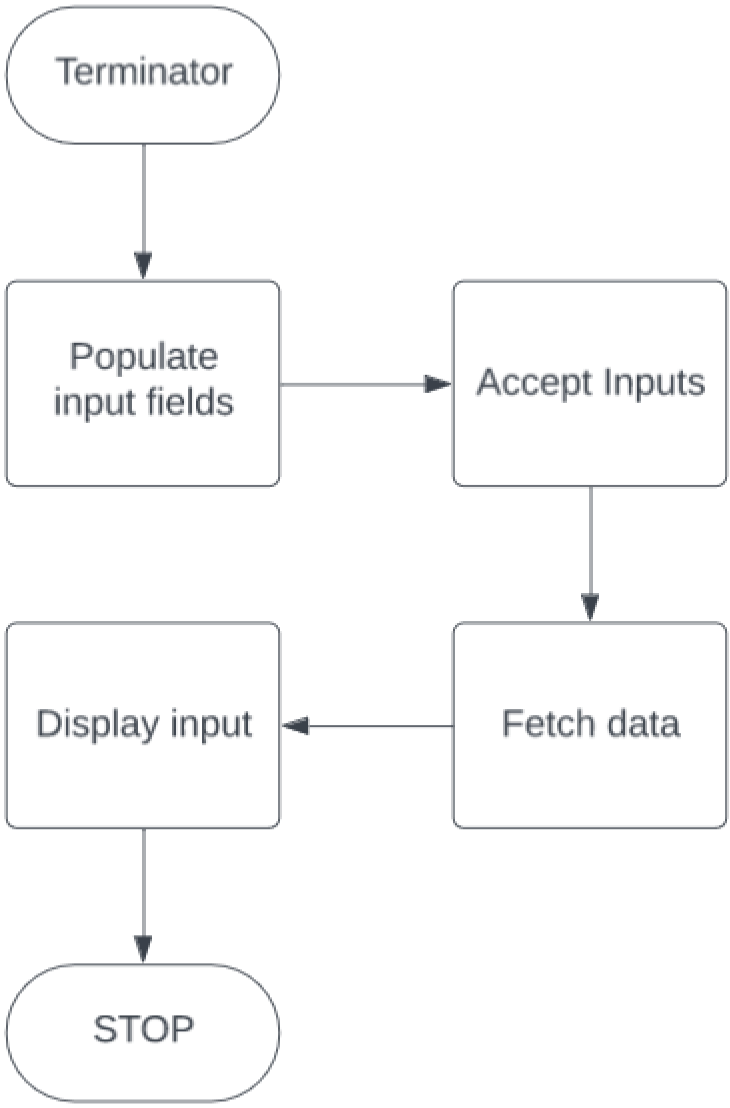


Figure 3.1.b Flowchart of project architecture - Front End

3.3. Algorithm and Process Design

MODEL

1. ***Data Collection***:
   Gather historical data on crop yields along with corresponding weather data, soil characteristics, and farming practices.

2. ***Data Preprocessing***:
   Clean the data to handle missing values, outliers, and inconsistencies. Normalise/standardise numerical variables if necessary.

3. ***Feature Selection***:
   Choose relevant features that could impact crop yield (example: temperature, rainfall, nutrient levels, etc.) for the regression model.

   For crop recommendation we're using datasets with the following features:
   - N:P:K ratio
   - temperature
   - humidity
   - pH
   - rainfall

   For yield prediction the dataset contains the mentioned features:
   - name of the crop
   - rainfall
   - pesticides (tonnes)
   - temperature

4. ***Model Selection***:
   Choose regression/classification algorithms based on the nature of the data and the problem. Common choices include linear regression, polynomial regression, decision tree regression, and support vector regression, as well as linear classifiers, K-nearest neighbours, etc.
   In our specific case we've selected the following models:

   - *Random Forest regressor*
   Random Forest is an ensemble learning method that combines the predictions from multiple decision trees. In many real-world scenarios, relationships between variables are non-linear. Random Forests can capture complex non-linear relationships as each tree can model different parts of the input space. Random Forest is less prone to overfitting compared to a single decision tree, especially when the dataset has many labels.

   - *SVM classifier*
   SVMs perform well in high-dimensional spaces. This makes them suitable for text classification, image recognition, and various other real-world problems where data can have many features. SVMs generally tend to have good generalisation performance and are less prone to overfitting, especially in high-dimensional spaces.

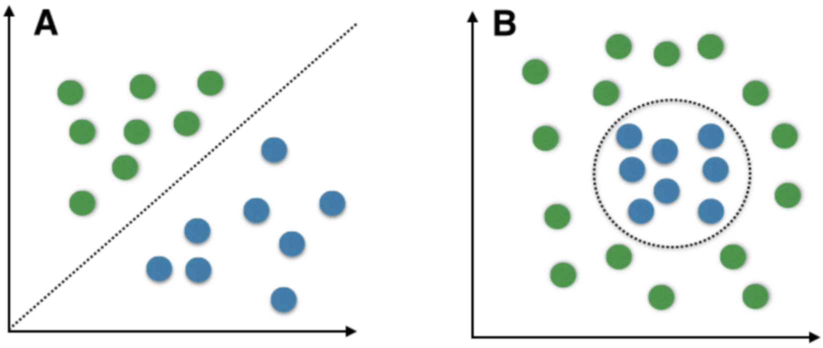


Figure 3.2. Visual representation of working of SVM

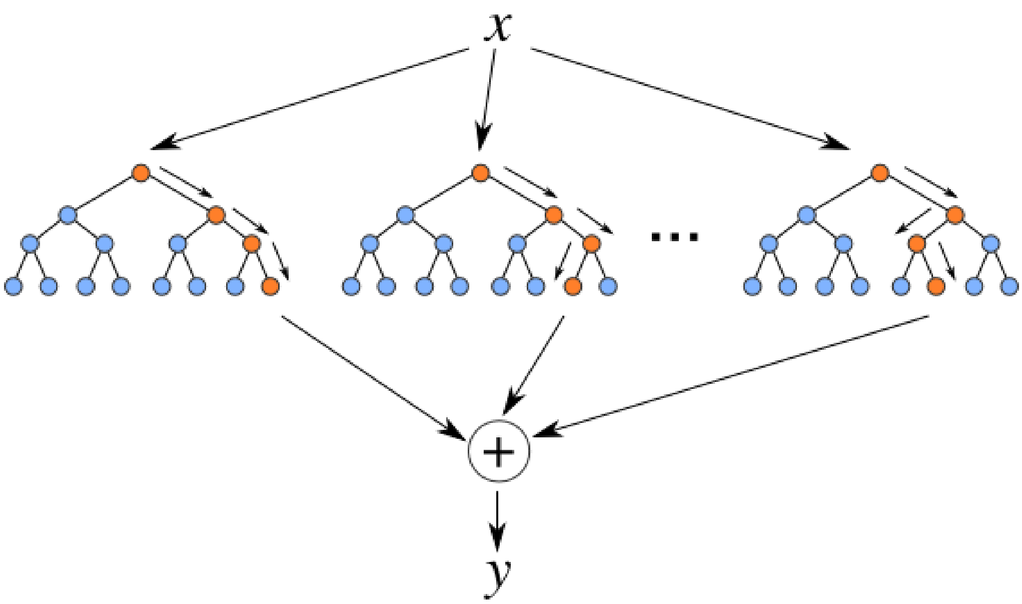


Figure 3.3. Visual representation of working of Random Forest

5. ***Model Training***:
   Split your data into training and testing sets. Train the models on the training data, allowing it to learn the relationships between the features and the crop yield.

6. ***Model Evaluation***:
   Use the testing data to evaluate the model's performance.

7. ***Interpretation***:
   Analyse the coefficients/features to understand which factors are most influential in predicting crop yields (and accordingly weight them).

8. ***Fine-tuning***:
   Adjust model parameters to improve its performance as per satisfaction.

FRONT END

1. ***Population***:
   Populate the page with inputs and options as per the selected features. Load sample dataset.

2. ***Input***:
   Wait for user input. Alternatively, autofill the input if commanded.

3. ***Fetch***:
   On form submission fetch data from the backend in the form of a prediction.

4. ***Display***:
   Show the prediction to the user.

3.4. Details of Hardware & Software

Hardware requirements: A mobile device or PC

Space required:
Only **556 KB** space (only **136 KB** without image and font metadata)
For hosting:
4GB RAM and 100MB space (ideally).

Software requirements:
Front-end:
Basic HTML5, SCSS/Sass, Vite for bundling, SCSS

Back-end: Python (Flask)

Database: MySQL (SQLite)

ML: Scikit-learn

Figure 3.4. Current tech stack

3.5. Experiment and Results

Results:

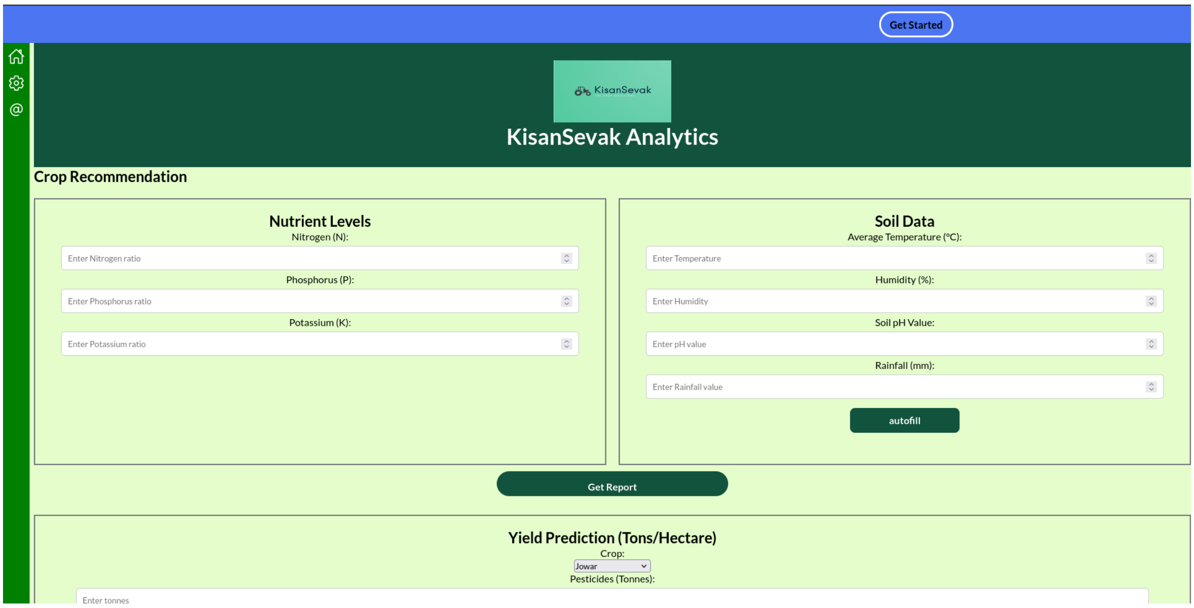


Figure 3.5.a Screenshot of layout

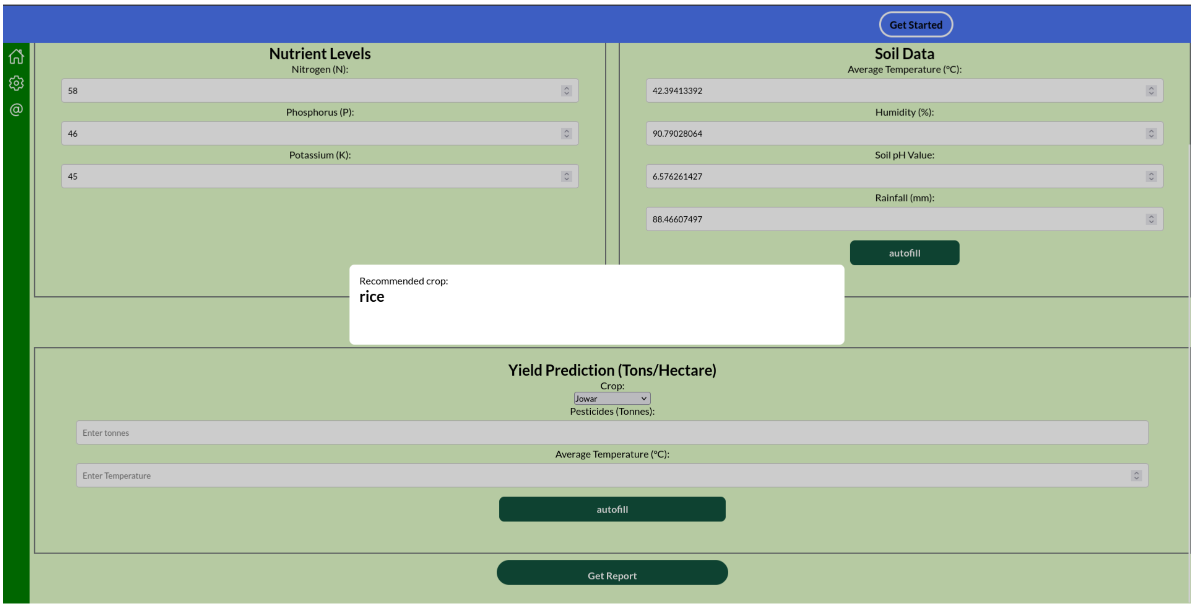


Figure 3.5.b Crop recommendation screenshot

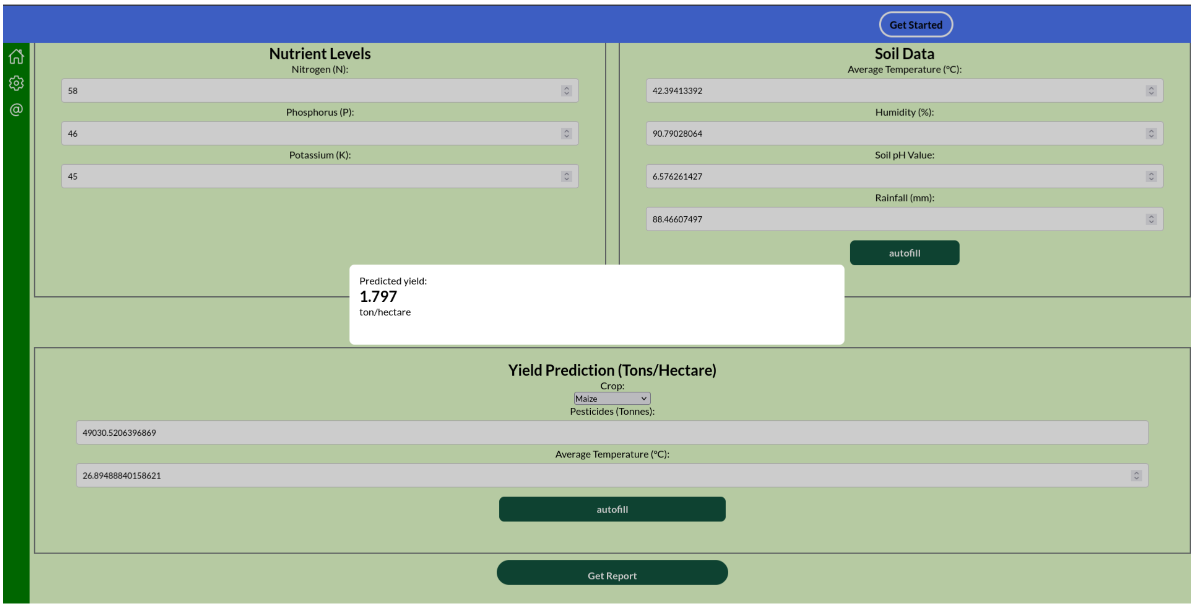


Figure 3.5.c Yield prediction screenshot

SUMMARY OF DATASETS USED:

This is a rough description of the datasets used.

	mean	std	min	25%	50%	75%	max
ton/ha yield	8.09	9.59	0.66	1.37	2.81	11.2471	38.58
pesticides tonnes	48459	14997	14485	37423	46195	61257	75000
avg_temp	26.01	0.90	23.2600	25.46	25.98	26.67	28.85

Table 3.1. Description of yield prediction dataset

	mean	std	min	25%	50%	75%	max
N	50.54	36.779	0	21	37	84	140
P	53.51	33.252	5	28	51	68	145
K	48.57	51.173	5	20	32	49	205
temperature	25.61	5.077	8.82	22.78	25.54	28.63	43.36
humidity	71.38	22.405	14.25	60.23	80.43	90	99.98
ph	6.46	0.784	3.50	5.95	6.42	6.93	9.93
rainfall	103.41	55.22	20.21	64.51	94.86	124.70	298.56
label	10.41	6.32	0	5	10	16	21

Table 3.2. Description of crop recommendation dataset

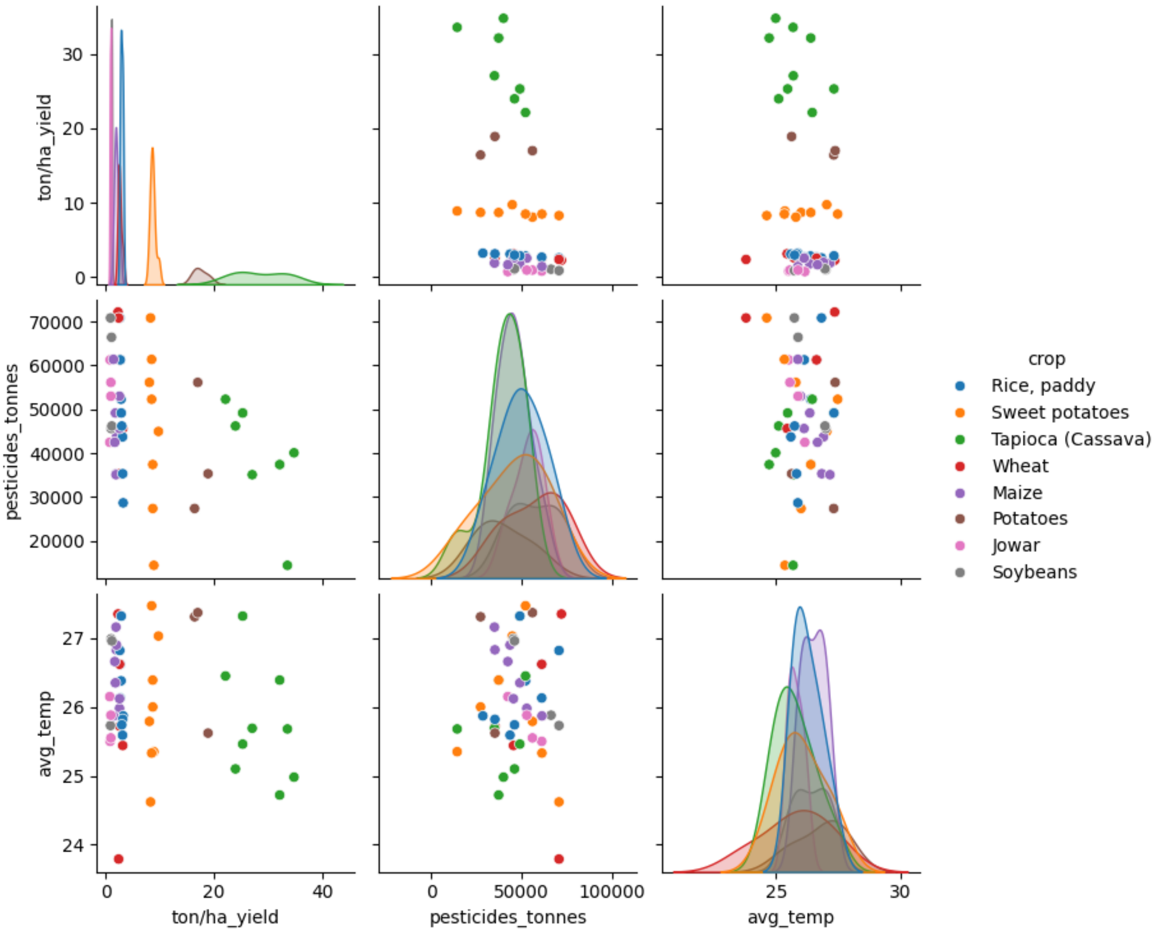


Figure 3.6. Plotting of sample data

Evaluation:

For the regression model the following measures were used to evaluate the efficiency.

- Mean Absolute Error (MAE): It provides a straightforward assessment of the model's accuracy.
- Root Mean Squared Error (RMSE): It penalises larger errors more heavily, giving a better understanding of the overall prediction accuracy.

We have achieved an RMSE of **5.47 x 10^{-4}** with the current Random Forest regression model.

For the classification model we have used the confusion matrix, precision, recall and F1 score as our metrics.
Given are the formulae for the same:

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n}$$

$$Precision = \frac{T_p}{T_p + F_p}$$

$$Recall = \frac{T_p}{T_p + T_n}$$

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

The significance of the given measures are as follows:

- **Precision:** If a model has a high precision, it means that it is very good at identifying positive cases. However, it may not be very good at identifying all of the positive cases (i.e., it may have a low recall).

- **Recall:** If a model has a high recall, it means that it is very good at identifying all of the positive cases. However, it may not be very good at distinguishing between positive and negative cases (i.e., it may have a low precision).

- **F1 score:** The F1 score is a good metric to use when both precision and recall are important. It is a harmonic mean of precision and recall, which means that it takes both metrics into account.

- **Accuracy:** Accuracy is a simple metric to understand, but it can be misleading in some cases. For example, if a dataset is imbalanced (i.e., there are many more negative cases than positive cases), a model can achieve a high accuracy by simply predicting the negative class for all cases.

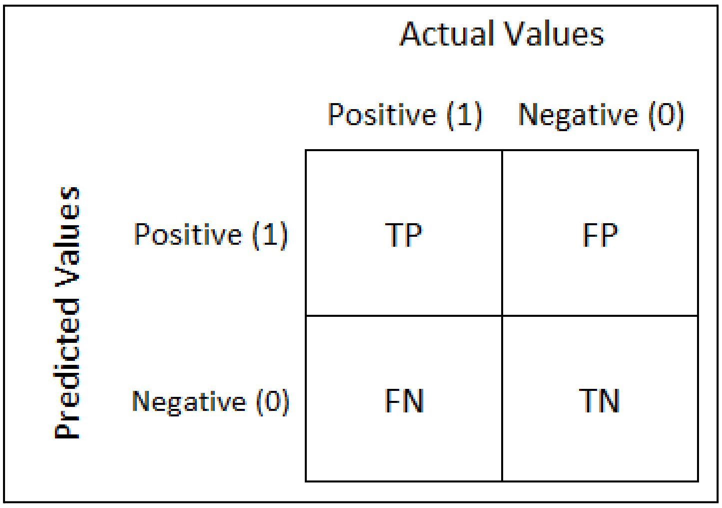


Figure 3.7. Confusion Matrix

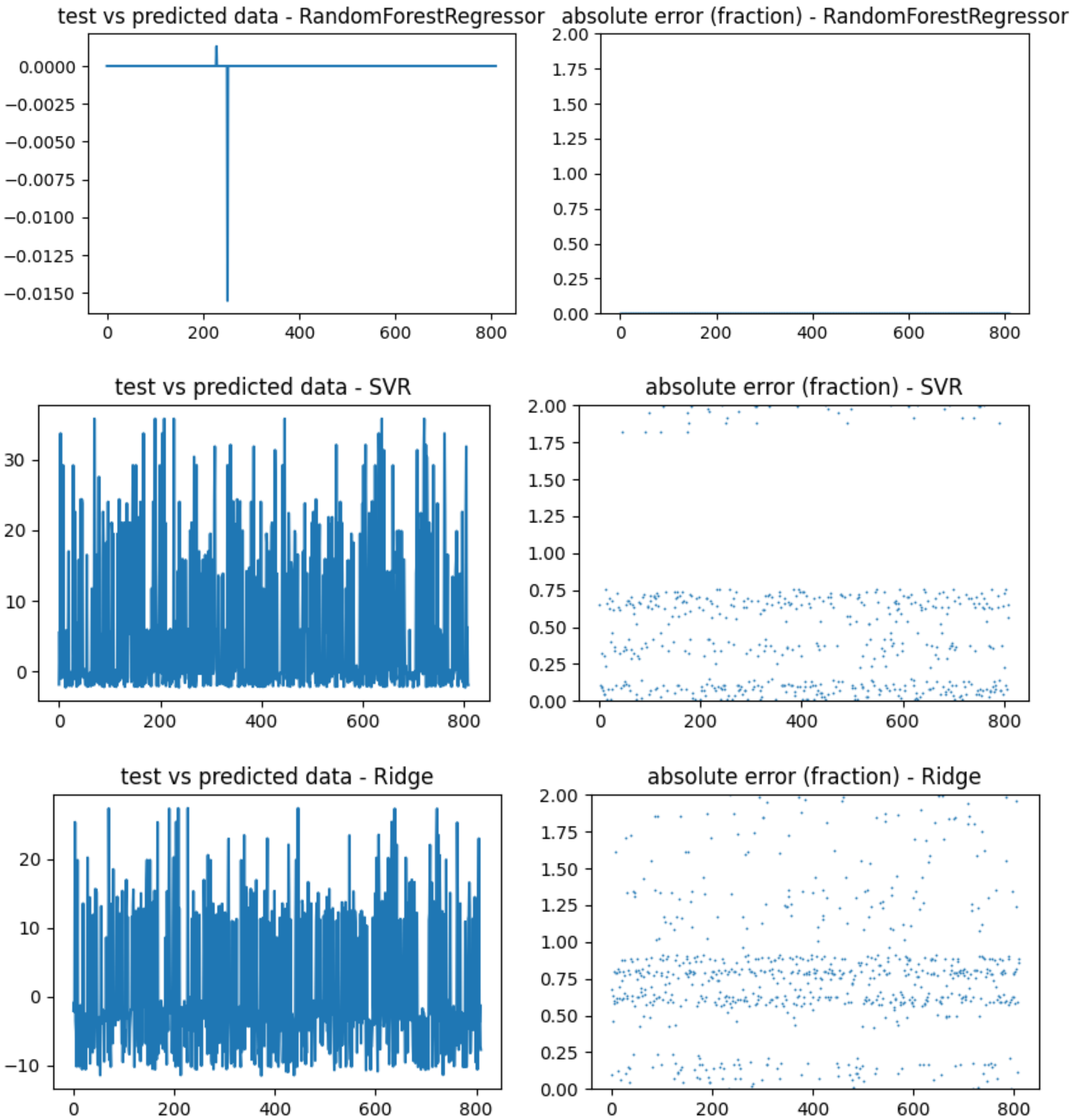


Figure 3.8. Drastic difference between evaluated models

The above set of graphs of the mean absolute errors of various models are shown in order to put into perspective the importance of selecting the right ML model. Random Forest gives a low absolute error due to the fact that it can handle non-linear relationships between features and the label much more accurately than most models. In our particular case the pesticide, temperature and crop features for yield prediction have a somewhat non-linear relationship.

3.6. Conclusion and Future work

The proposed digital platform offers a transformative solution to the challenges faced by Indian farmers in the agriculture sector. By leveraging advanced technologies, it empowers farmers with informed decision-making tools, direct market access, and knowledge resources. The platform's holistic approach aims to break the cycle of chronic indebtedness, improve agricultural practices, and foster a more equitable and prosperous future for farmers and the nation. Successful implementation requires collaboration and has the potential to reshape Indian agriculture, ultimately benefiting both individual farmers and the broader agricultural landscape.

Future work:

- Addition of various graphs and simulation of *real time* data.
- Better datasets and feature engineering to generate more versatile models.
- Market price prediction for various crops.
- Localization by adding local language support.
- Addition of shopping module.

References

Research papers referred:

- Agriculture Overview: From production-centric to farmer-centric - VisionIAS Weekly Focus - *March 2022*

- Comparing agroecosystems: Effects of cropping and tillage patterns on soil, water, energy use and productivity - Cambridge University Press - *September 2004*

- Artificial intelligence in the agri-food sector - STOA - *January 2023*

- Seeds of Silicon: Internet of Things for Smallholder Agriculture - MIT D-Lab - *2019*

- P. Malik, S. Sengupta and J. S. Jadon, "Comparative Analysis of Soil Properties to Predict Fertility and Crop Yield using Machine Learning Algorithms," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 2021, pp. 1004-1007, doi: 10.1109/Confluence51648.2021.9377147.

- A. Shah, R. Agarwal and B. Baranidharan, "Crop Yield Prediction Using Remote Sensing and Meteorological Data," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 952-960, doi: 10.1109/ICAIS50930.2021.9395849.

- K. M P and D. N R, "Crop Prediction Based on Influencing Parameters for Different States in India- The Data Mining Approach," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2021, pp. 1785-1791, doi: 10.1109/ICICCS51141.2021.9432247.

- Li, Y., Chen, J., Drury, C.F. et al. The role of conservation agriculture practices in mitigating N2O emissions: A meta-analysis. Agron. Sustain. Dev. 43, 63 (2023).

Other references:

- Economic Survey of Maharashtra - Government of Maharashtra - *2021-2022*
https://bankofmaharashtra.in/writereaddata/documentlibrary/74de861a-a869-47a0-b1f7-456f32e976ee.pdf
- Economic Survey of Maharashtra 2021-22
http://mls.org.in/pdf2022/budget/ESM_2021_22/Economic%20Survey%20of%20%20Maharashtra%202021-22.pdf