PRECISION FARMING USING MACHINE LEARNING AND DATA ANALYTICS

PROJECT MENTOR:
MR. M. SUNDARABABU

312419104001 - AASHIK MATHEW P
312419104025 - BHARATHWAJ N

INTRODUCTION

- Farming is a remarkable part of the economy in India, as it adds 18% of the absolute GDP.
- Farming is extremely vulnerable to climate change. Higher temperature eventually reduce yields of desirable crops while encouraging weed and pest proliferation.
- Risks for agriculture in developing countries include production risks like drought and floods that destroy crops in the field.
- Food production decreasing, food price rising and land fertility decreasing results in agriculture land decreasing.
- Precision farming provides a way for the farmers to find out the appropriate crop for plantation by considering various climate factors.

ABSTRACT

- Precision farming, a web application prototype for a real-time crop recommendation using Machine Learning and Data Analytics.
- The business logic of python uses Machine Learning techniques in order to predict the most profitable crop in the forecasted weather and soil conditions at a specified location.
- The datasets are obtained from Kaggle. Datasets of other countries can also be utilized for the project.
- The server is coded in order to access the web application outside the client's browser.
- The server code once executed will navigate to the displayed IP Address on the prompt to access the web application.

LITERATURE SURVEY

S.NO	YEAR	TITLE	DESCRIPTION	ADVANTAGE	DISADVANTAGE
1	2022	Precision agriculture using IoT data analytics and machine learning	ML is applied using IoT data to agriculture to increase the quality and quantity of production	By using the real-time data for prediction the accuracy and usability will be high.	Setting up of IoT devices is costly
2	2021	Towards paddy rice smart farming: a review on big data, machine learning, and rice production tasks	Emerging technologies enable new abilities for prediction	The data collected from sensors influences ML algorithms	high cost, potential privacy and security risks, need for technical expertise, over-reliance on technology,
3	2021	Precision agriculture using advanced technology of IoT, unmanned aerial vehicle, augmented reality, and machine learning	PA uses site-specific crop management concept based on measured data to find the cause of yield reduction.	Using the data collected from real-time sensors will enable to predict the cause for yield reduction precisely.	It is costly to setup IoT devices for prediction purposes

S.NO	YEAR	TITLE	DESCRIPTION	ADVANTAGE	DISADVANTAGE
4	2020	Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications	The performance of those models highly relies on the quality of the extracted features	High complexity, high computational demands, sensitivity to input data quality, and potential for overfitting	Unable to create a direct linear mapping between the raw data and crop yield values
5	2021	Towards precision agriculture: IoT-enabled intelligent irrigation systems using deep learning neural network	It provides solution to deliver bigger, better, and more profitable yields with fewer resources	By collecting the soil and atmosphere data using IoT devices the irrigation required for the field can be predicted precisely.	Setting up of IoT devices is costly
6	2021	An efficient IoT based smart farming system using machine learning algorithms	It uses efficient prediction method called WPART based on machine learning techniques to predict crop productivity	WPART based on machine learning techniques is used to predict the crop productivity.	It is costly to setup IoT devices for prediction purposes
7	2019	Machine learning techniques in wireless sensor network based precision agriculture	Advancements in IoT, Wireless Sensor Networks makes it easier to predict	Wireless sensor is used for collecting the data required for prediction which makes is accessible from places.	The number of interconnected devices increases, the load also increases.

S.NO	YEAR	TITLE	DESCRIPTION	ADVANTAGE	DISADVANTAGE
8	2019	A smart agriculture IoT system based on deep reinforcement learning	The presented system Integrates artificial intelligence for smart agriculture	By using the real-time data for prediction the accuracy and usability will be high.	Setting up of IoT devices is costly
9	2020	Machine learning applications for precision agriculture: A comprehensive review	Precision agriculture has emerged as an innovative tool to address current challenges in agricultural sustainability.	It uses new technologies for increasing the quality and quantity of production in agriculture	Dataset has to be highly reliable in order to automate the agriculture process. Such dataset require high cost for setup real-time.
10	2019	Data-driven decision making in precision agriculture: the rise of big data in agricultural systems	It employs the practice of big data analysis to solve problems which reveal promising areas of use.	The implementation is challenged by its high complexity and volume of the data produced.	The high volume and complexity of the data produced pose challenges in successfully implementing.
11	2020	Smart farming becomes even smarter with deep learning - A bibliographical analysis	Advancements in connectivity, automation, and AI enable farmers better to apply precise treatments determined by machines.	It explains the concepts of smart farming clearly and those activities carried out by data scientists.	It does not explain the need and necessity of a very reliable dataset for prediction.

S.NO	YEAR	TITLE	DESCRIPTION	ADVANTAGE	DISADVANTAGE
12	2018	Deep learning models for the prediction of rainfall	Deep Learning enabled the self-learning data labels which allows to create a data-driven model for a time series dataset	It allows to make change detection from the time series data and also predicts the future event's	Requires large amount of training data and computing power.
13	2020	Application and prospective discussion of machine learning for the management of dairy farms	It shows how ML methods have already been used and describe the potential that these may offer	ML methods were applied to predict data in a variety of areas in dairy farming	The implementation found not to be reliable upon tested algorithms
14	2018	Machine learning application on agricultural datasets for smart farm enhancements	It aims to show how to manage heterogeneous information and data coming from real datasets.	It uses multiple machine learning algorithms and chooses the accurate one for prediction	Limited exploitation of IoT and need for skilled workforce and knowledge investment.
15	2021	Smart crop prediction using IoT and Machine Learning	It uses IoT for real-time monitoring the field and using the humidity, temperature, soil data for prediction.	IoT based crop prediction improves agriculture by monitoring field real-time.	Potential inaccuracies in data analysis and crop recommendations, leading to potential loss in crop output and quality.

EXISTING SYSTEM

- Majority of the precision farming methods are IoT based data acquiring systems for prediction which requires capital investment for setup in each and every region. It is quite impossible as there are a lot of region that has to be covered and the investment required for this setup is costly.
- Current existing system uses crop dataset, soil dataset which are acquired from kaggle and Government websites. Weather dataset is acquired from IoT device using Arduino.
- Using real-time dataset makes the prediction more reliable for real time agriculture. They use Naive Bayes algorithm for prediction model which is then deployed in a mobile application.
- However reliable the real-time dataset is, a detailed dataset is required for ML to analyze the pattern of the data. IoT device which is collecting real time data can be more reliable in future after 5-10 years, but not in a basic stage of creating a new dataset now at present. It is advisable to use the pre-existing dataset for weather.
- On the other hand it is necessary to train and test in different ML algorithms and choosing the most accurate one for prediction which they failed to do. These are the major drawbacks in the existing system.

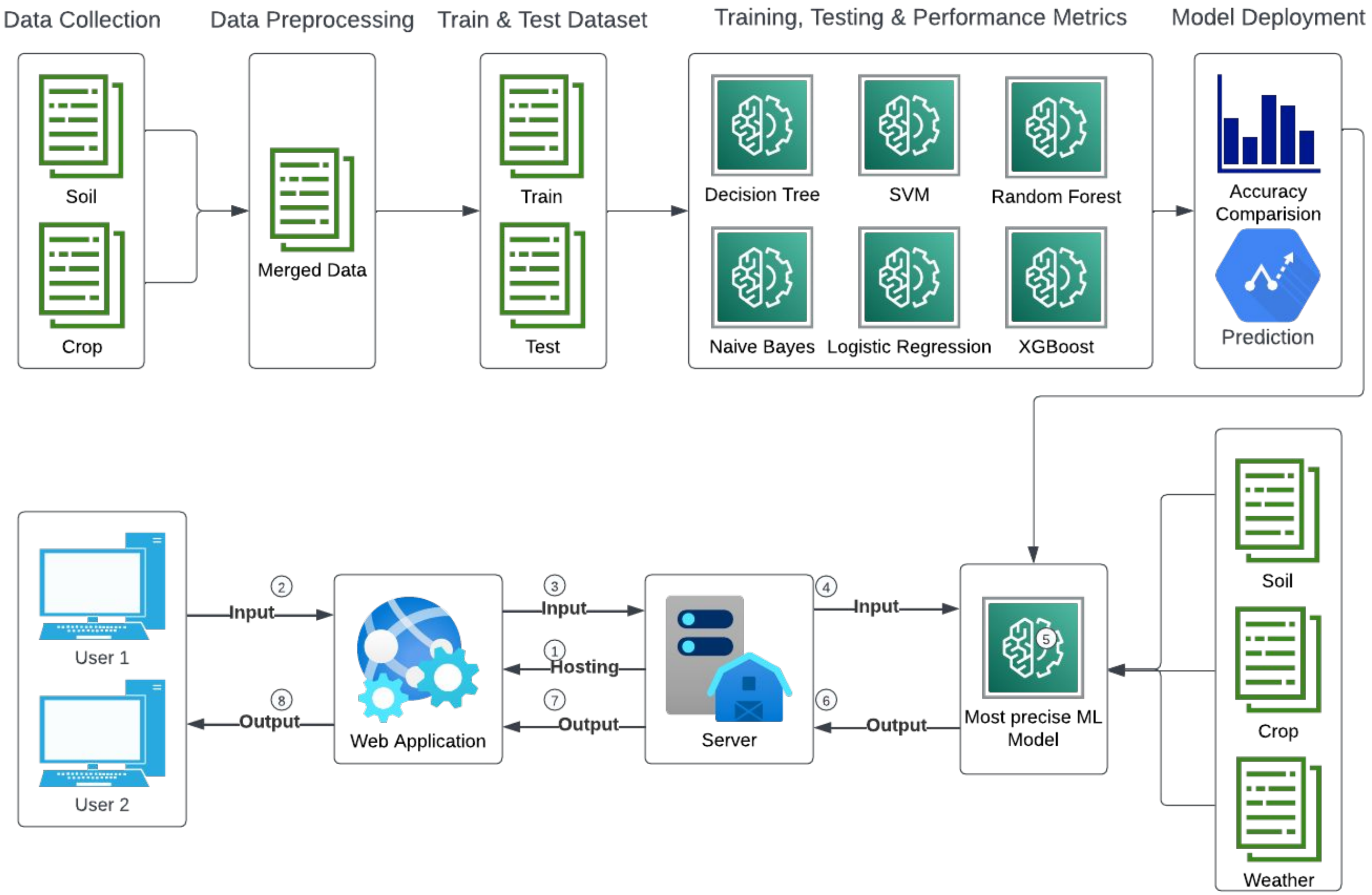
PROPOSED SYSTEM

- The proposed system will integrate the data obtained from the soil, crop repository during data integration process and the weather department dataset will be accessed using an API.
- Different ML algorithm are used and the model is trained using the integrated dataset. 30% of the dataset is used for training and 70% is used for testing.
- The accuracy of all ML models is compared by plotting in a bar chart and the most accurate model is used in the web application for user prediction.
- We use machine learning libraries and frameworks such as pandas, numpy, scikit-learn to optimize the model, fine-tune its parameters, and prevent overfitting or underfitting.
- Web application is developed using HTML, CSS and JavaScript for users to access this application worldwide.
- The server is deployed in Flask Environment.

PROPOSED SYSTEM ADVANTAGE

- Using multiple Machine Learning algorithms and comparing them will give us a precise idea of the trustworthiness of the prediction system.
- Apart from the crop and soil dataset, accessing a weather dataset using API makes it even more easier and developer friendly.
- The web application provides accessibility to people of different regions around the world.
- The appropriate crop prediction by analyzing the environmental factors will increase the crops productivity and maintains the GDP of India.
- It helps controlling insects, pests and soil borne diseases thereby minimizing the land infertility issues.
- Precision farming can help farmers adopt sustainable practices by reducing the environmental impact of agriculture. For example, precision farming can help farmers conserve water, reduce chemical runoff, and minimize greenhouse gas emissions.

SYSTEM ARCHITECTURE



MODULES

1. **Data collection:** This module involves collecting data from a variety of sources, such as weather stations, soil sensors, crop repository to create a comprehensive dataset for analysis.
2. **Data preprocessing:** This module involves cleaning and integrating the data for analysis, including tasks such as missing value imputation, outlier detection, and feature scaling.
3. **Model training:** This module involves splitting the processed data into training and testing data. The split up data is then passed on through various ML Algorithms (Decision Tree, Naive Bayes, SVM, Logistic Regression, Random Forest and XGBoost) for training and testing.
4. **Model testing:** This module involves using test data and evaluating the performance of the trained model, by calculating precision, recall, F1 Score and support.
5. **Model Deployment:** This module involves comparing the accuracy of various ML Models to choose the most accurate model for deployment.

DATA COLLECTION

- **Datasets:** Crop dataset, Soil dataset and Weather dataset.
- The crop and soil datasets are acquired from kaggle. These datasets are chosen as contains the list of crops that satisfies the climatic conditions and soil conditions required for its grow.
- The weather dataset is accessed using API in the state and city selection options in the UI.
- **Crop dataset:** It contains 31 types of crops along with a 100 variations of climatic conditions that are suitable for its growth. The climatic conditions includes temperature, humidity, pH, rainfall.
- **Soil dataset:** It contains 97 types of crops along with a 19 variations of soil conditions that are suitable for its growth. The soil conditions includes Nitrogen (N), Phosphorous (P), Potassium (K), pH.
- **Weather dataset:** It contains all the states and cities of India. It contains the humidity and temperature data of the chosen region. This dataset is accessed using WeatherAPI.

DATA PRE-PROCESSING

1. **Data Cleaning:**

- The case of the dataset is converted into lowercase and few labels are renamed.
- The crop and soil dataset has a common parameter between them which is the **‘crop label’**. This can be used to segregate the common data in both the datasets.
- The unique crop names of crop dataset is stored in **‘crop_names’** and from soil dataset is stored in **‘crop_names_from_soil’**.
- The **‘crop_names_from_soil’** is compared with **‘crop_names’**. By doing so the common crop names in both the datasets are stored in **‘extract_labels’**.
- The common crop names are chosen in both the crop and soil dataset, which are then stored in two new CSV files **‘crop_merged’** and **‘soil_merged’**. This new RAW Dataset contains the crops that have common **‘crop label’** between them.

2. Data Integration:

- The **‘soil_merged’** dataset is described to understand the count, mean, SD, minimum and maximum.
- The Nitrogen (N), Phosphorous (P), Potassium (K) columns are plotted for visual understanding.
- The data of **‘soil_merged’** is copied in **‘reco_soil’** to create NPK values.
- ‘**Temp**’ is created containing NPK columns to create values in range of crop names in **‘crop_merged**’.
- The N value of crop is obtained from soil dataset and a random value is generated between -20 and 20.
- The P value of crop is obtained from soil dataset and a random value is generated between -5 and 20.
- The K value of crop is obtained from soil dataset and a random value is generated between -5 and 5.
- NPK Column is created in **‘crop_merged’** and values are merged from **‘Temp’.**
- The integrated crop and soil dataset is saved as **‘crop_recommendation’** containing crop, temperature, humidity, pH, rainfall, N, P and K values.

MODEL TRAINING

- The integrated dataset **'crop_recommendation'** is used for training the model.
- The unique and count of **'crop label'** of the dataset is listed.
- The dataset is separated into features and targets where N, P, K, temperature, humidity, pH, rainfall are features and crop label is the target.
- The dataset is split into training and test data. 20% of the dataset is used for training and 80% for testing.
- XTrain, YTrain is used for training the model.
- Multiple Machine Learning algorithms are used for training. They are Decision Tree, Logistic Regression, Naive Bayes Classifier, Random Forest, SVM Classifier, XGBoost.
- The model once trained is then tested using test data and the model is saved in pickle format .**pkl**

MODEL TESTING

- After training the model using XTrain and YTrain, the model is tested using XTest and YTest.
- The test data contains only the features which has to be used to predict the targets.
- By comparing the result of XTest and YTest the test accuracy is calculated.
- The accuracy calculation involves precision, recall, F1 Score and support.
- Cross validation is also carried out for all the models so that it generalizes well to new data.
- Empty **‘accuracy’** and **‘model_lists’** are created to store the accuracy values of corresponding models.
- The calculated accuracy is added to the **‘accuracy’** list and the corresponding model name is added to the **‘model_lists’**.
- For each and every model after training the testing and test accuracy calculation process is carried out.
- After training and testing, the model is saved in **.pkl** format.

ACCURACY CALCULATION

- Precision : (True Positives) / (True Positives + False Positives)
- Recall : (True Positives) / (True Positives + False Negatives)
- F1 Score : 2 * (Precision * Recall) / (Precision + Recall)
- Support : Number of samples of the true response that lies in each class
- Accuracy : (True Positives + True Negatives) / (Total Samples)

➔ **True Positives:** When you predict an observation belongs to a class and it actually does belong to a class.

➔ **True Negatives:** When you predict an observation does not belong to a class and it actually does not belong to a class.

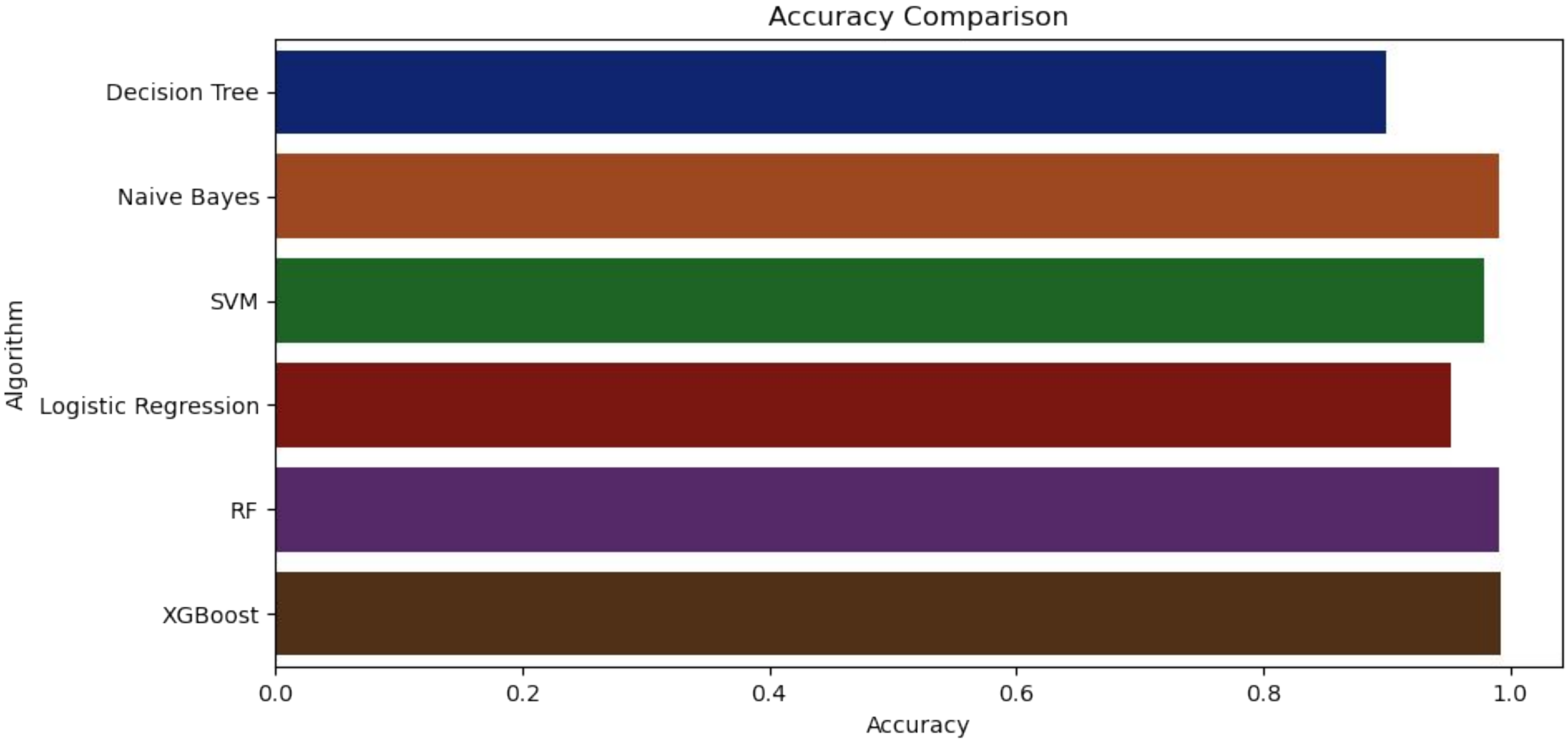
➔ **False Positives:** When you predict an observation belongs to a class when in reality it does not.

➔ **False Negatives:** When you predict an observation does not belong ot a class when in fact it does.

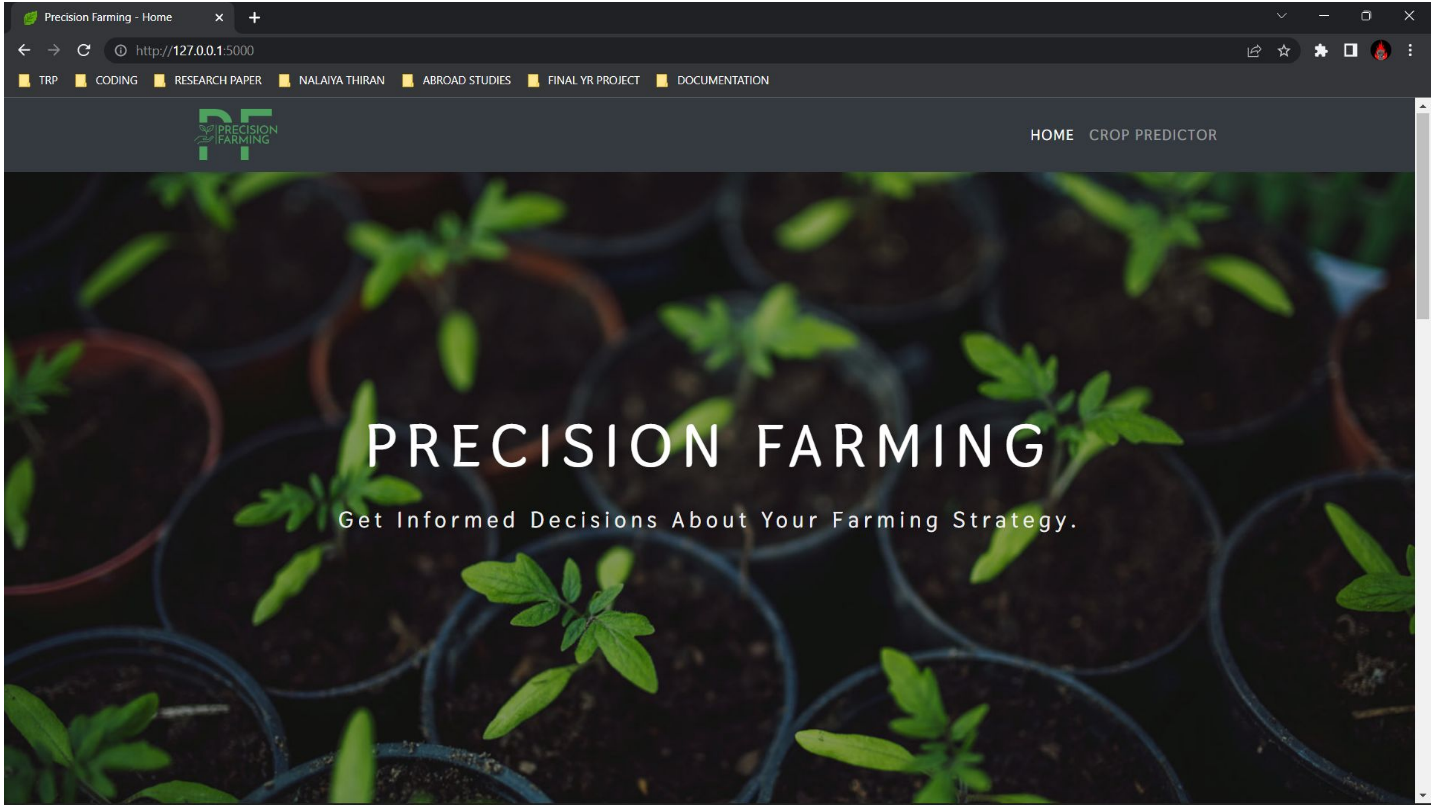
MODEL DEPLOYMENT

- Once all the six models are trained, tested and accuracy is calculated all the accuracy and model names would have been stored in **‘accuracy’** and **‘model_lists’.**
- The accuracy of all the models is plotted in bar plot for a detailed understanding of its accuracy and to choose the highly accurate model for prediction.
- **‘accuracy’** is provided as the X Label for the bar plot.
- **‘model_lists’** is provided as the Y Label for the bar plot.
- In case if all the accuracy are similar with slight deviations it cannot be easily identified by plotting. To solve such circumstances the model name and accuracy is also printed in the console.
- The most accurate model **(XGBoost / Random Forest)** is used for prediction and model deployment.

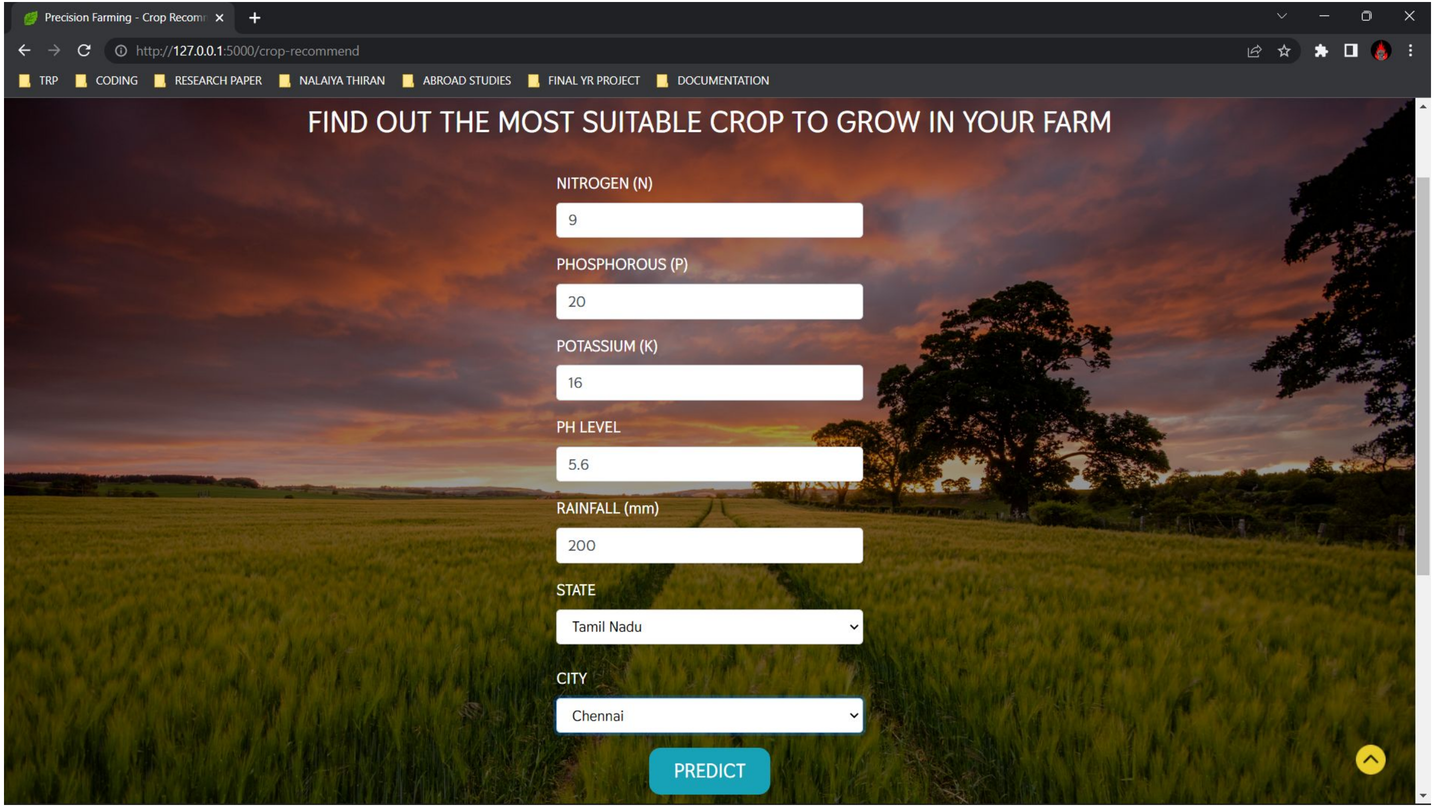
MODEL COMPARISON



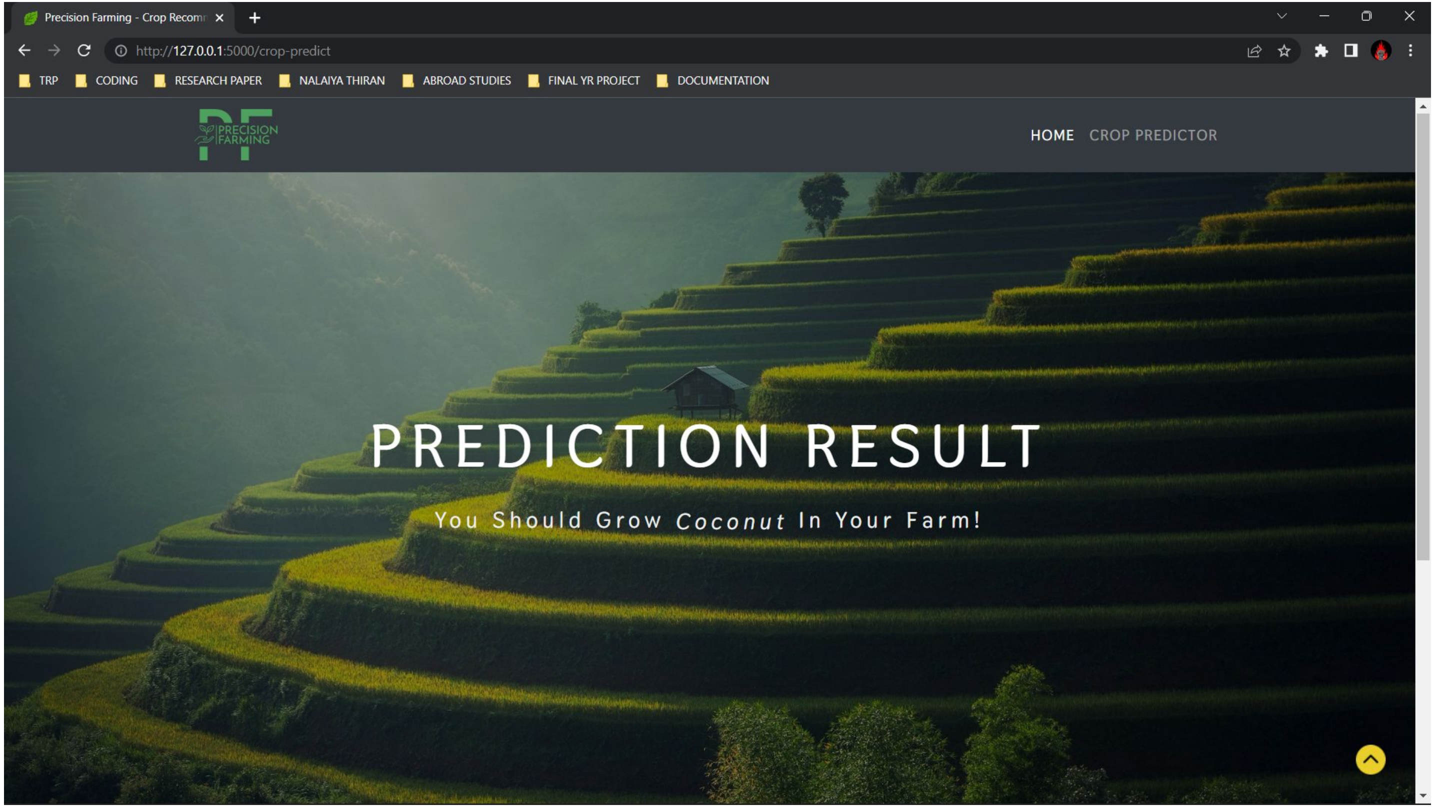
OUTPUT SCREENSHOTS



OUTPUT SCREENSHOTS



OUTPUT SCREENSHOTS



PERFORMANCE METRICS

S.NO	ALGORITHM	TRAINING ACCURACY	TESTING ACCURACY
1	Decision Tree	88.18%	90%
2	Naive Bayes	99.60%	99.09%
3	Support Vector Machine (SVM)	99.54%	97.95%
4	Logistic Regression	97.38%	95.22%
5	Random Forest	100%	99.09%
6	**XGBoost**	**100%**	**99.31%**

PERFORMANCE METRICS

S.NO	CONTENTS	EXISTING SYSTEM	PROPOSED SYSTEM
1	Data Collection	Weather - IoT Device Crop, Soil - Kaggle	Weather - **API** Crop, Soil - Kaggle
2	Machine Learning Algorithm	Naive Bayes	**XGBoost** / Random Forest
3	Training Data	30%	30%
4	Testing Data	70%	70%
5	Accuracy	97.00%	**99.31%**
6	Application Type	Mobile Application	Web Application

CONCLUSION

- Predicting the appropriate crop for farming by considering the environmental factors results in increased crop yield.
- Chances for drought and floods will be considered and the crop which could sustain such climatic conditions will be predicted.
- Good production of crops will result in increase of the economic GDP of India and those countries which uses this web application.
- It helps the farmers make informed economic and management decisions but also support famine prevention efforts.

FUTURE WORK

- This project can be improvised by implementing crop yield prediction during the process of selection of suitable crops.
- By doing so, the farmers will have the knowledge of the suitable crop that has the highest and lowest yield considering the environmental factors.
- Multiple Linear Regression algorithm can be used for this improvisation. Crop yield dataset is required for its implementation.
- Apart from these, crop irrigation prediction can also be implemented to this project. It provides the information of periodical irrigation required for the plantation of crops considering the environmental factors such as temperature, humidity, etc.

REFERENCES

[1]. Akhter, Ravesa, and Shabir Ahmad Sofi. "Precision agriculture using IoT data analytics and machine learning." *Journal of King Saud University-Computer and Information Sciences* 34.8 (2022): 5602-5618.

[2]. Alfred, Rayner, et al. "Towards paddy rice smart farming: a review on big data, machine learning, and rice production tasks." *IEEE Access* 9 (2021): 50358-50380.

[3]. Ponnusamy, Vijayakumar, and Sowmya Natarajan. "Precision agriculture using advanced technology of IoT, unmanned aerial vehicle, augmented reality, and machine learning." *Smart Sensors for Industrial Internet of Things*. Springer, Cham, 2021. 207-229.

[4]. Kashyap, Pankaj Kumar, et al. "Towards precision agriculture: IoT-enabled intelligent irrigation systems using deep learning neural network." *IEEE Sensors Journal* 21.16 (2021): 17479-17491.

[5]. Rezk, Nermeen Gamal, et al. "An efficient IoT based smart farming system using machine learning algorithms." *Multimedia Tools and Applications* 80.1 (2021): 773-797.

[6]. Mekonnen, Yemeserach, et al. "Machine learning techniques in wireless sensor network based precision agriculture." *Journal of the Electrochemical Society* 167.3 (2019): 037522.

[7]. Bu, Fanyu, and Xin Wang. "A smart agriculture IoT system based on deep reinforcement learning." *Future Generation Computer Systems* 99 (2019): 500-507.

[8]. Sharma, Abhinav, et al. "Machine learning applications for precision agriculture: A comprehensive review." *IEEE Access* 9 (2020): 4843-4873.

[9].Tantalaki, Nicoleta, Stavros Souravlas, and Manos Roumeliotis. "Data-driven decision making in precision agriculture: the rise of big data in agricultural systems." *Journal of agricultural & food information* 20.4 (2019): 344-380.

[10]. Zeynep Unal, *et al.* “Smart farming becomes even smarter with deep learning - A bibliographical analysis”, IEEE Access (Volume 8), pp 2169-3565, 2020

[11]. S Aswin, P Geetha, R Vinayakumar, *et al.* “Deep learning models for the prediction of rainfall”, IEEE Conference on Communication and Signal Processing, pp 657-661, 2018

[12].Marianne Cockburn *et al.* “Application and prospective discussion of machine learning for the management of dairy farms”, 10(9), 2020

[13]. Fabrizio Balducci, Donato Impedovo, *et al.* “Machine learning application on agricultural datasets for smart farm enhancements:, 6(3), 2018

[14]. Dhivya Elavarasan, P. M. Durairaj Vincent. *et al.* “Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications”, IEEE Access (Volume 8), pp. 86886-86901, 2020

[15].Archana Gupta, Dharmil Nagda, *et al.* “Smart crop prediction using IoT and machine learning”, International Journal of Engineering Research & Technology, 2021

THANK YOU!