Precision Farming – A Machine Learning Based Crop Recommender

Dr. K. Madan Mohan[1], M. Rochani Reddy[2], M. Mani Kumar[3], G. Ravi Teja[4]

[1]Associate Professor, Department of Computer Science & Engineering, Guru Nanak Institute of Technology
[2,3,4]Student, Department of Computer Science & Engineering, Guru Nanak Institute of Technology

---***---

ABSTRACT: In rural India, agriculture and its related industries are without a doubt the main sources of income. The nation's Gross Domestic Product (GDP) is also significantly influenced by the agriculture sector. The vast scale of the agriculture industry is a blessing for the nation. The yield per acre of crops, however, is disappointing when compared to global norms. This is among the potential reasons why India's marginal farmers commit suicide at a higher rate. A practical and easy-to-use yield forecast method for farmers is suggested in this research. Farmers can connect to the planned system through a smartphone application. GPS aids in determining the user's location. The user enters the soil type and region. Predicting the crop yield for a user-selected crop or selecting the most lucrative crop list are made possible by machine learning algorithms. K-Nearest Neighbor (KNN), Random Forest (RF), Support Vector Machine (SVM), and Multivariate Linear Regression (MLR) are some of the machine learning methods used to forecast agricultural productivity. With a 95% accuracy rate, the Random Forest produced the best results out of all of them. In order to increase yield, the system also recommends the optimal time to apply fertilizers.

Key Words: Precision Farming, Machine Learning, Crop Recommendation, Farm Data Analysis, Soil Quality Assessment

I. INTRODUCTION

In India, agriculture has a long history. India is currently the world's second-largest producer of agricultural products. About half of the workforce was employed in agriculture-related sectors like forestry and fishing, which accounted for 16.6% of the 2009 GDP. India's agricultural sector is contributing less money to the country's GDP [1]. The main element influencing agricultural finances is crop yield. Numerous factors, including meteorological, geographic, organic, and economical aspects, influence crop yield [6]. Because market prices fluctuate, farmers find it challenging to determine which crops to plant and when [7]. According to Wikipedia statistics, for the past ten years, India's suicide rate has been between 1.4 and 1.8% per 100,000 people. Because of the unpredictability of climatic circumstances, farmers are unsure of which crop to cultivate and when and where to begin. Changes in seasonal climates and essential resources like soil, water, and air also make it questionable how different fertilizers should be used. The agricultural yield rate is continuously dropping in this situation [2].

Providing farmers with an intelligent and easy-to-use recommender system can effectively address challenges in agriculture. One of the most significant difficulties in this field is accurately predicting crop yield [3]. Farmers often rely on their previous experiences [4]and knowledge of specific crops to estimate how much they will produce [3]. Several key factors, such as weather conditions, pest infestations, and harvest planning, play a crucial role in determining agricultural output. Therefore, having precise and reliable information about past crop performance is essential for making informed decisions and managing risks in farming [5].

II. LITERATURE REVIEW

Umamaheswari S [1], Sreeram S, Kritika N, Prasanth DJ. The fundamental promise of blockchain technology for the agriculture sector is that it eliminates the requirement for third parties that are typically needed to maintain confidence in buyer-seller relationships—or, for that matter, in any interaction between a source and a destination. Block chain technology makes it possible for transactions to occur peer-to-peer without the need for middlemen. In addition to facilitating peer-to-peer transactions, blockchain technology may generate "smart contracts," which carry out the terms of any agreement in response to certain triggers. Every time value is transferred, whether it be in the form of cash, goods, or services, the transaction can be recorded, establishing a permanent record of the item or transaction from its origin to its final destination. In this field, blockchain technology can be quite beneficial. By placing all of the data on agricultural occurrences on a blockchain, a transparent and reliable system can be created. On a single platform, farmers may also obtain real-time information about soil moisture, demand and sale price, payments, climate and environment-related data, and seed quality. The project's goal is to store sensor data in a blockchain and create a smart contract that can be used to buy and sell land and crops on the Ethereum network.

According to Jain A [2], ensuring stable agricultural growth has been a persistent challenge in India. To examine the fluctuations in rice production, the study evaluates 41 years of data (from 1970–71 to 2011–12), focusing on the area cultivated, total production, and yield of paddy. The analysis reveals that, at the overall level of India, the compound annual growth rate of area, production, and yield of rice was positive, but it had been gradually declining over the periods. In the most recent decade (2000–01 to 2011–12), there has been an increase in instability. The low proportion of irrigated land to total cropped land, as well as the decrease in the usage of seeds, manure, and other agricultural inputs, could be the causes of the rise in instability. During the post-reform period (1990–1991 to 2016–2017), the wholesale prices of paddy have shown increased volatility across different states, whereas the prices received by farmers at harvest have exhibited comparatively lower fluctuations.

Manjula E, Djodiltachoumy S [3]. A new area of study in crop production analysis is data mining. Predicting yield is a crucial problem in agriculture. Every farmer wants to know how much harvest he may anticipate. In the past, farmers' experiences with a given crop and field were taken into account when predicting production. One significant problem that needs to be resolved with the data at hand is the yield forecast. For this, data mining techniques are a superior option. In agriculture, several data mining techniques are employed and assessed to predict crop production for the upcoming year. This study introduces and implements a technique for forecasting crop yields using historical agricultural data. The approach involves applying association rule mining to uncover patterns within the dataset. The primary objective is to create a predictive model that can be utilized for estimating future crop yields effectively. This research presents a concise study on predicting crop yields using a data mining approach based on association rules, specifically applied to a selected region in Tamil Nadu, India. The experimental results indicate that the proposed technique provides accurate and reliable crop yield predictions.

Sagar BM, Cauvery NK [4]. Because it provides for basic needs, agriculture is essential to human survival. It is common knowledge that the vast majority of Indians (≥55%) work in agriculture. There are obstacles to raising crop production in India because of differences in the country's climate. Reaching intended goals in agri-based crop productivity has become a difficult task. A number of factors that directly affect agricultural productivity and production must be taken into account, including climate, geography, economics, and politics. Predicting crop output is one of the key components of agricultural operations. To increase crop yield, farmers must have knowledge about crop yield before planting seeds in their farms. Data analytics is one such trend that has permeated the agricultural industry and is utilised for crop yield management and crop health monitoring. In recent years, the adoption of technology in agriculture has significantly increased. Advances in the sector have highlighted the value of big data, making stakeholders more aware of its potential. However, a major challenge lies in evaluating the effectiveness and overall impact of big data analytics when implemented in agricultural practices.

Research is being done to find out how big data analytics may be applied to increase agricultural practices' production. Crop yield forecast, crop health monitoring, and other related operations are aided by the study of agricultural data. Numerous research about the application of data analytics in the field of agriculture can be found in the literature. The current study provides information on a range of data analytics techniques used to forecast agricultural yield. Additionally, the work highlights significant gaps in the suggested field of study.

Wolfert S, Ge L, Verdouw C, Bogaardt MJ [5]. "Smart farming" refers to the advancement that highlights the integration of information and communication technologies within the cyber-physical cycle of farm management. Emerging technologies such as cloud computing and the Internet of Things (IoT) are expected to build upon this progress, leading to a greater integration of robotics and artificial intelligence in agriculture. The term "Big Data" describes the extensive and varied sets of data that can be gathered, analyzed, and leveraged to support informed decision-making in the farming sector. The purpose of this review is to learn more about the most recent developments in big data applications for smart farming and to pinpoint the socioeconomic issues that need to be resolved. A conceptual framework for analysis that may be applied to further research on this subject was created using an organised methodology. The study indicates that the use of big data in smart farming impacts not just the early phases of crop production but also extends its effects throughout the entire food supply chain..

III.METHODOLOGY

It is assumed that the data owner is reliable, and the users accessing the data have received proper authorization from the data owner. The communication channels between the owner and users are secure on existing security protocols such as SSL, TLS. With regard to the cloud server, our scheme resists a more challenging security model which is beyond the "semi-honest server" used in other secure semantic searching schemes. In the proposed model, the untrusted cloud server may try to provide incorrect or falsified search results and gain access to sensitive data. However, it is assumed that the server does not intentionally delete or alter the stored documents. As a result, the secure semantic framework must ensure both data confidentiality and result verifiability within this security context.

Existing System Disadvantages

- The primary obstacle encountered in the agricultural industry is ignorance of the shifting climate. Every crop has unique climatic characteristics that make it suited. Using precision farming techniques can help with this. Precision farming raises the production yield rate in addition to maintaining agricultural productivity.
- The current crop yield recommendation system is either hardware-based and expensive to operate, or it is not readily available.
- There are still unresolved issues with developing a user-friendly crop suggestion tool, even with the several solutions that have lately been put out.

PROPOSED SYSTEM:

We have presented a model in this project that tackles the current problems. The suggested system's unique feature is that it suggests the most lucrative crop for the area while also assisting farmers in maximising crop production. In order to maximise crop output and hence help satisfy the nation's growing demand for food supply, the suggested model offers crop selection based on economic and environmental factors. Through the analysis of variables like rainfall, temperature, area, season, soil type, etc., the suggested model forecasts crop yield. Additionally, the technique aids in determining when fertiliser should be applied.

Proposed System Advantages

- The suggested model forecasts the crop yield based on the region's data sets. By boosting yields and optimising the resources used, combining agriculture with machine learning will lead to further improvements in the agricultural sector. The primary components for predicting current performance are statistics from prior years.
- The suggested technique suggests the best time to apply fertiliser using a recommender system.
- The suggested system's techniques include boosting crop productivity, analysing crops in real time, choosing effective parameters, making better decisions, and achieving higher yield.

SYSTEM ARCHITECTURE

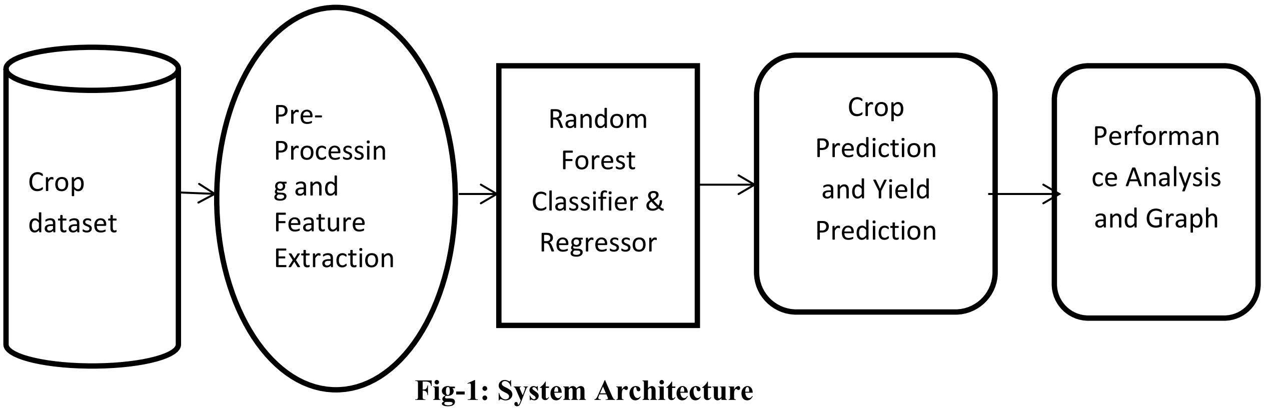

Fig-1: System Architecture

It begins with data pre-processing, where raw agricultural data is cleaned and transformed to make it suitable for model training. The processed dataset is then utilized to train a Random Forest Classifier. Once training is complete, the model predicts the most appropriate crop by considering input parameters like soil nutrient levels, temperature, and humidity. To validate the model's effectiveness, its performance is assessed using relevant evaluation metrics to ensure accurate and dependable crop recommendations.

MODULES:

1. **Data Collection:**

Data collection is the initial and essential phase in building a machine learning model. The quality and quantity of the collected data directly influence the model's overall performance—the more comprehensive and accurate the data, the more effective the model will be. Various techniques can be employed to gather data, such as manual collection, web scraping, and others. For this project, the dataset utilized is named crop_yield.csv.

2. **Dataset:**

There are 94375 distinct data points in the dataset. The dataset consists of six columns.

- State Name
- Season
- Crop
- Area
- Production
- Soil Types
- Data Preparation

The data will be transformed. by eliminating certain columns and missing data. We will start by compiling a list of the column names that we wish to preserve.After that, we eliminate or drop every column save for the ones we choose to keep. Lastly, we eliminate from the data set the rows that have missing values.

3. **Model Selection:**

Two datasets are required when building a machine learning model: one for testing and one for training. We only have one currently, though. Let's divide this in half using an 80:20 ratio. Additionally, we will separate the dataframe into feature and label columns. Here, we imported the sklearn train_test_split function. Next, divide the dataset using it. Additionally, it splits the dataset with 80% as the train dataset and 20% as the test dataset when test_size = 0.2. A random number generator that aids in dataset splitting is seeded by the random_state option. Four datasets are returned by the function. The data was divided into four parts labeled as test_x, test_y, train_x, and train_y. The shape of these sets shows how the dataset is split. To build the model, we utilized the Random Forest Classifier, which fits multiple decision trees to the training data. The model was trained by applying the fit method with train_x and train_y as inputs. After training, the model's performance is evaluated by using the predict method on the test_x data.

Analyze and Prediction:

We selected just four features from the real dataset:

State Name: India is a union of states and union territories for administrative purposes.

Season:

*Winter,which lasts from December through February.

*The pre-monsoon or summer season, which runs from March to May.

*The rainy season, also known as the monsoon or kharif crops, which run from June to September.

*The fall or post-monsoon season, which runs from October to November.

*Around the middle of November, the rabi crops are planted.

Types of Soil: Area of various soil types: how much space? They gathered

4. **Precision on the test set**:

On the test set, our accuracy was 0.95%

IV. IMPLEMENTATION

The crop recommendation system was developed using Python with Spyder IDE, leveraging the Random Forest algorithm to predict most suitable crops and ideal fertilizer application time based on inputs such as soil type, geographical region and season. The system utilized a cleaned dataset containing over 94,000 entries from states like Maharashtra and Karnataka, featuring attributes like crop type, cultivated area, production, and season. After the preprocessing, the data was splitted into training and testing sets in a 80:20 ratio. The Random Forest model achieved an accuracy of 95%. A user-friendly web interface was built with HTML and was created to enable farmers to enter the data and receive tailored crop suggestions. The entire system was developed on a Windows machine equipped with an Intel i7 processor, 1TB HDD and 16GB RAM.

V. EXPERIMENTAL RESULTS

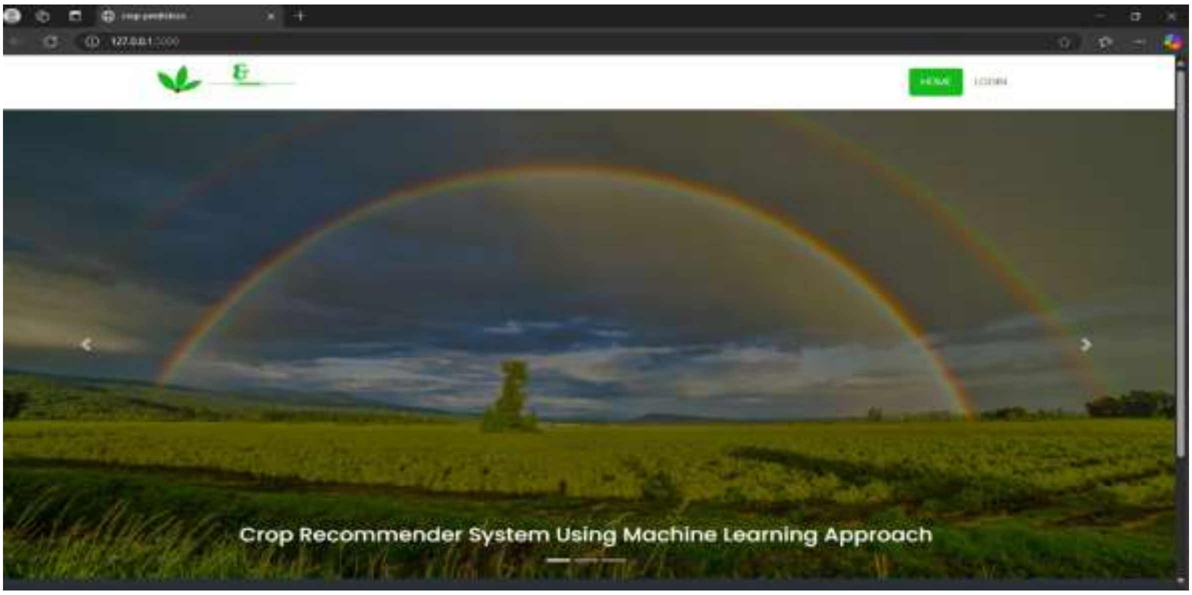

Fig-2: Home Page of the Crop Recommender System Web Application

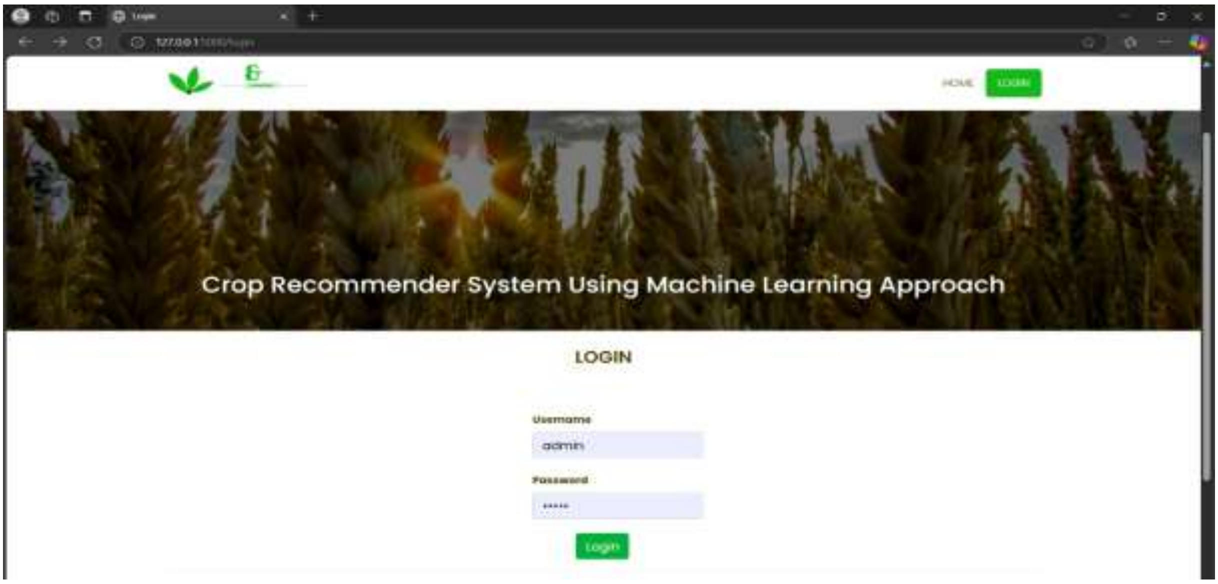

Fig-3: Login Page of the Crop Recommender System Web Application

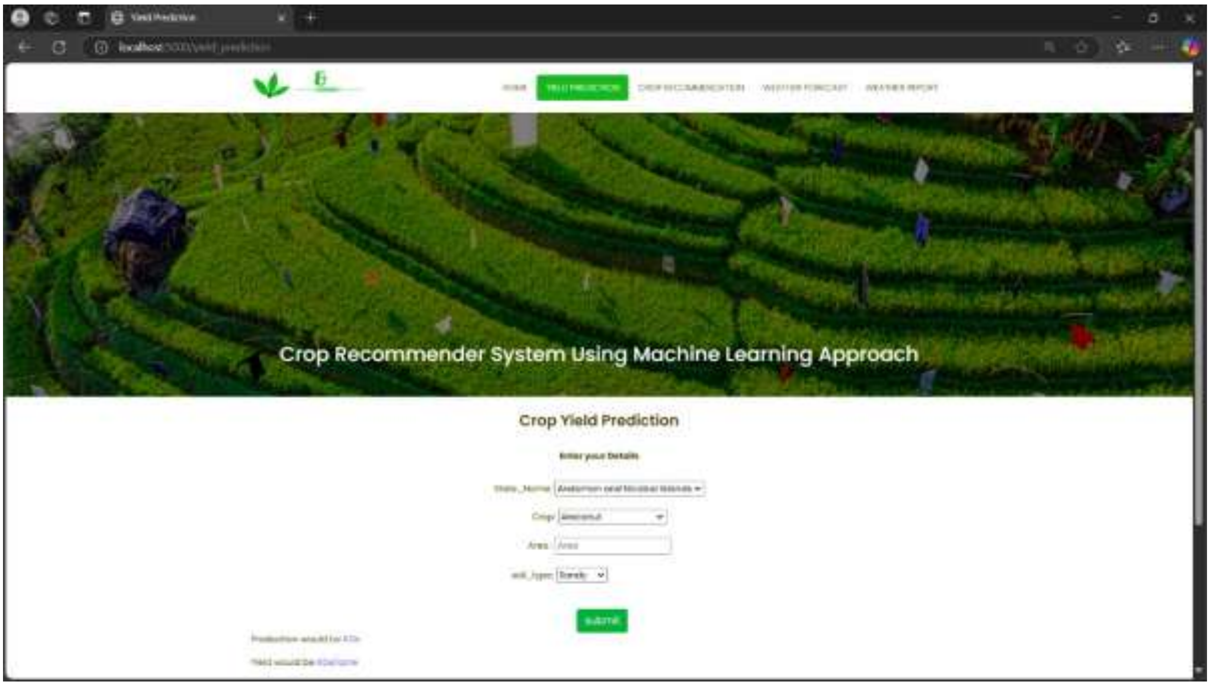

Fig-4: File Upload Interface of the Crop Recommender System

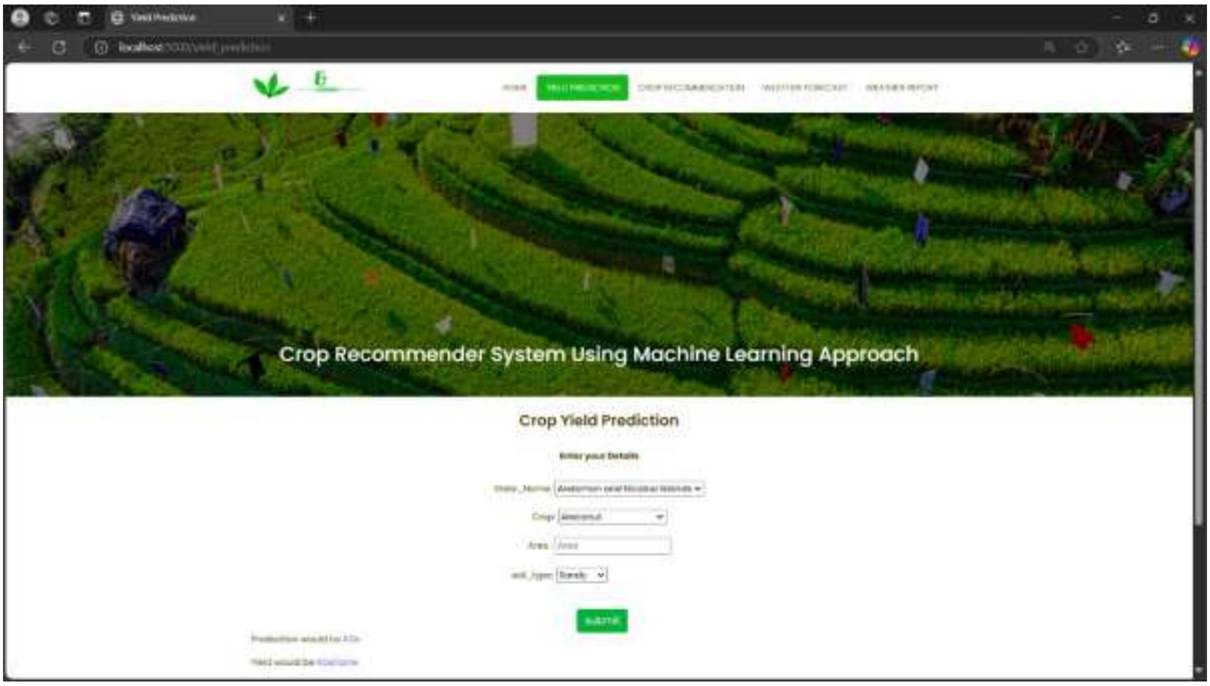

Fig-5: Crop Yield Prediction Interface of the Crop Recommender System

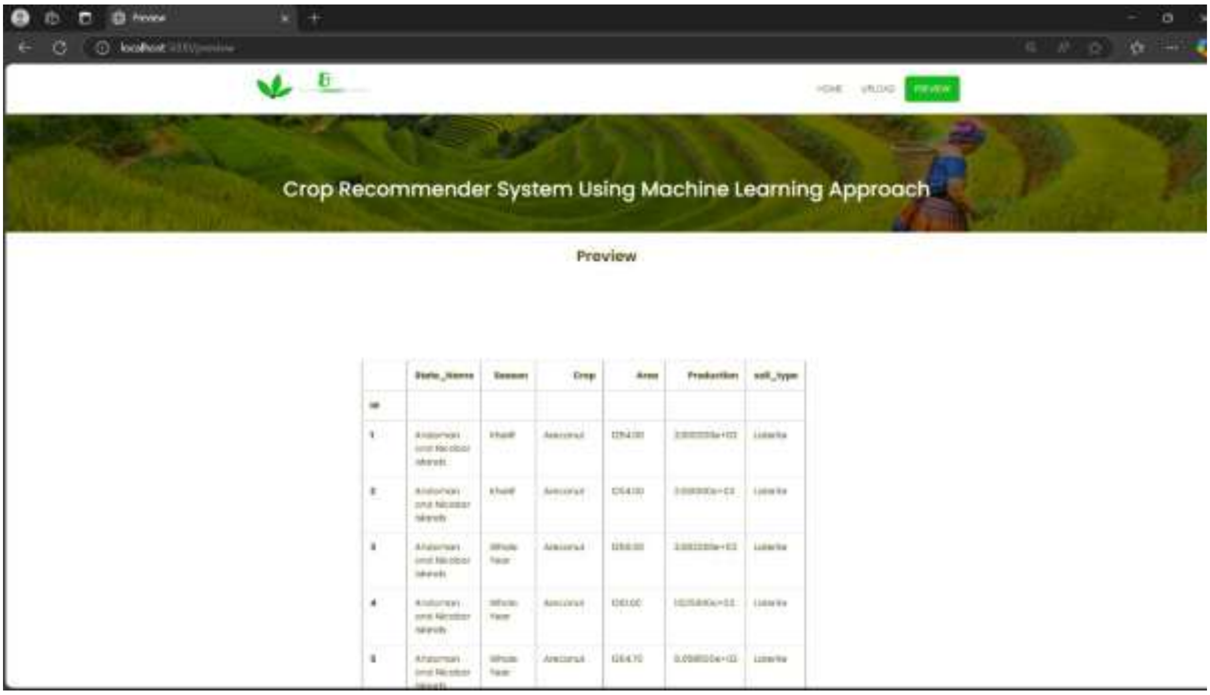

Fig-6: Dataset Preview Page of the Crop Recommender System

Fig-7: Visualization of Crop and Soil Data in the Crop Recommender System

VI. CONCLUSION

This article focused on the limits of existing technology and their applicability for yield prediction. Next, it gives farmers a demonstration of a practical yield forecasting system after that; the recommended approach connects them. The integrated predictor system allows farmers to predict the production of a crop. The integrated recommender system allows the user to investigate possible crops and their yields to help them make better decisions. To ascertain yield to accuracy, machine learning methods were implemented and assessed on the given datasets. Additionally, the proposed model looked at the best time window for applying fertiliser.

VII. FUTURE ENHANCEMENT

Future research will concentrate on periodically updating the databases to generate precise forecasts, and the procedures can be automated. Providing the appropriate fertiliser type for the crop and region is another feature that needs to be put into place. To put this into practice, a comprehensive analysis of the available fertilisers and how they relate to soil and climate must be conducted. It is necessary to analyse the statistical data that is already available.

REFERENCES

[1] Umamaheswari S, Sreeram S, Kritika N, Prasanth DJ, "BIoT: Blockchain-based IoT for Agriculture", 11th International Conference on Advanced Computing (ICoAC), 2019 Dec 18 (pp. 324-327). IEEE.

[2] Jain A. "Analysis of growth and instability in the area, production, yield, and price of rice in India", Journal of Social Change and Development,2018;2:46-66

[3] Manjula E, Djodiltachoumy S, "A model for prediction of crop yield" International Journal of Computational Intelligence and Informatics,2017 Mar;6(4):2349-6363.

[4] Sagar BM, Cauvery NK., "Agriculture Data Analytics in Crop Yield Estimation: A Critical Review", Indonesian Journal of Electrical Engineering and Computer Science, 2018 Dec;12(3):1087-93.

[5] Wolfert S, Ge L, Verdouw C, Bogaardt MJ, "Big data in smart farming–a review. Agricultural Systems", 2017 May 1;153:69-80.

[6] Jones JW, Antle JM, Basso B, Boote KJ, Conant RT, Foster I, GodfrayHC, Herrero M, Howitt RE, Janssen S, Keating BA, "Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. Agricultural systems", 2017 Jul1;155:269-88.

[7] Johnson LK, Bloom JD, Dunning RD, Gunter CC, Boyette MD, Creamer NG, "Farmer harvest decisions and vegetable loss in primary production. Agricultural Systems", 2019 Nov 1;176:102672.