CROP COMBINATION AND MARKET PRICE PREDICTION

CS6611 – CREATIVE AND INNOVATIVE PROJECT

Submitted by

Mathankumar P	**(2022103002)**
Sanjay T G	**(2022103507)**
Gurumoorthi R	**(2022103525)**
Sri balaji J	**(2022103536)**

in partial fulfilment of the requirements for the award of the degree of

BACHELOR OF ENGINEERING

in

COMPUTER SCIENCE AND ENGINEERING

College of Engineering, Guindy

ANNA UNIVERSITY, CHENNAI - 600 025

APRIL 2025

ANNA UNIVERSITY, CHENNAI - 600 025

BONAFIDE CERTIFICATE

Certificate that this project request titled **Crop combination and Market Price Prediction** is the Bonafide work of **Mathankumar P (2022103002), Sanjay TG (2022103507), Gurumoorthi R (2022103525)** and **Sri balaji J (2022103526)** who carried out the project work under my supervision, for the fulfillment of the requirements as part of the CS6611 – Creative and Innovative Project.

Dr. V. MARY ANITA RAJAM

Professor & head

HEAD OF THE DEPARTMENT

Department of Computer Science and Engineering, College of Engineering Guindy, Anna University, Chennai – 600025.

Dr. P.UMA MAHESWARI

Professor

SUPERVISOR

Department of Computer Science and Engineering, College of Engineering Guindy, Anna University, Chennai – 600025.

Dr. P. GANESH KUMAR

Associate Professor

SUPERVISOR

Department of Computer Science and Engineering, College of Engineering Guindy, Anna University, Chennai – 600025.

ABSTRACT

In the real time of agriculture, one of the persistent challenges faced by farmers is the difficulty in selecting the right crop to grow based on varying soil conditions and weather patterns, coupled with the unpredictability of market prices. To address this issue, prior research has focused on crop prediction systems, laying the foundation for advancements in agricultural decision-making processes. Building upon this existing work, our proposed system aims to revolutionize crop selection and market price forecasting through the integration of advanced machine learning and statistical modeling techniques.

Our solution seeks to assist farmers in making informed decisions by recommending the main crop and suitable sub crop based on key soil parameters (N, P, K, pH) and weather data (humidity, rainfall, temperature). By harnessing the power of machine learning algorithms, specifically the Random Forest Classifier for crop prediction, and utilizing the SARIMA model for market price forecasting, our system aims to accurately predict which crops will thrive in specific conditions and what their market prices are likely to be.

This integration of cutting-edge technology not only enhances agricultural decision-making but also improves the economic outcomes for farmers by reducing uncertainty and optimizing crop planning. By leveraging state-of-the-art models, our system strives to achieve high accuracy in crop and sub crop recommendations as well as market price predictions, ultimately empowering farmers with actionable insights.

The benefits of this technological advancement extend to both farmers and agricultural stakeholders. Farmers stand to gain from optimized crop selection, leading to higher yields and better market prices. Additionally, the system offers predictive market insights, reducing the risk of financial losses due to price fluctuations. By providing data-driven solutions, our system enhances overall agricultural productivity and sustainability, fostering greater confidence and profitability for farmers.

ACKNOWLEDGEMENT

Foremost, we would like to express our sincere gratitude to our project guide, **Dr. P. UMA MAHESWARI**, Professor and **Dr. P. GANESH KUMAR**, Associate Professor, Department of Computer Science and Engineering, College of Engineering Guindy, Chennai for their constant source of inspiration. We thank them for the continuous support and guidance which was instrumental in taking the project to successful completion.

We are grateful to **Dr. S. MARY ANITA RAJAM**, Professor and Head, Department of Computer Science and Engineering, College of Engineering Guindy, Chennai for her support and for providing necessary facilities to carry out for the project.

We would also like to thank our friends and family for their encouragement and continued support. We would also like to thank the Almighty for giving us the moral strength to accomplish our task.

Mathankumar P **Sanjay T G** **Gurumoorthi R** **Sri Balaji J**

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER – 1
INTRODUCTION

The agricultural industry is undergoing a transformation driven by the advancement of data science and machine learning technologies. Traditionally, farmers rely on their experience and intuition when selecting crops and predicting market prices. However, the increasing availability of agricultural data presents an opportunity to make more data-driven decisions. According to recent reports, the agricultural sector continues to face challenges like unpredictable weather patterns, soil variability, and volatile market prices, which directly affect farmers' productivity and revenue. Our **Crop Combination and Market Price Prediction System** aims to address these challenges by leveraging data analytics and machine learning to help farmers make informed decisions.

Farmers often struggle to choose the right crop based on varying soil conditions and weather patterns. Additionally, market price fluctuations can lead to financial losses. This system recommends the most suitable crop and sub crop based on critical parameters and predicts their future market prices using historical data from **AGMARKNET** which is the project of Directorate of Marketing and Inspection, Department of Agriculture Cooperation & Farmers Welfare Ministry of Agriculture & Farmers Welfare, Government of India.

The **Crop Combination and Market Price Prediction System** integrates advanced machine learning techniques, including **Random Forest** Classifiers for crop prediction and **SARIMA** models for market price forecasting, to enhance the agricultural decision-making process. This system helps farmers reduce uncertainty, optimize crop selection, and predict market prices, ultimately leading to improved agricultural practices and financial stability.

CHAPTER 2
LITERATURE SURVEY

1.Crop Recommendation & Prediction

Jain, S., *et al.* [1] proposed a crop recommendation model by combining Random Forest and Recurrent Neural Networks (RNNs) based on historical climatic data recorded in Telangana. The Random Forest model, using dependent agro-climatic soil data, is not dynamic; the choice of which crop to recommend is static when soil is sampled. RNNs capture real-time weather forecasts. Farmers are presented with basic analysis that will help their decisions on the right crop type to grow given its current environmental setting, taking challenges of unpredictable weather into account. The output presentation delivers sowing period recommendations and irrigation suggestions; note, optimization options as well. The model proved valid performance; comparatively matching recommended crop types based on regional climatic modelling outputs. However, the approach does not consider farmers' economic limitations which fundamentally shape the decision-making process facing farmers when growing a crop; soil nutrient data presented in terms of nitrogen (N), phosphorus (P), and potassium (K) beyond physical soil characteristics and labor is another aspect that should have been included to enhance agronomic rationality.

Modi, D., *et al.* [2] implemented a crop recommendation model that analyzes Support Vector Machine (SVM) as a classification model that generates crops based on soil parameters, such as N, P, K values, rainfall, pH, humidity, and temperature. Their data set consisted of a complete range of soil parameters from distinct regions, which was a strong basis for classification. The model produced a high classification accuracy of 97% after completing sampling and predicting suitable crops in respect to soil and nutrients. The output for the final recommendation specifically attributes the right crop recommendation for farmers, improving productivity of agricultural workflow associated with land conditions. While their system performed with a high classification accuracy, the recommendation system did not use market price tendencies or dynamically update weather conditions, which is essential for real-time farming decisions. Furthermore, the omission of an economic component or multiple seasons of cropping was a significant limitation in its application.

Anjana, A. *et al.* [3] proposed a crop prediction algorithm based on weather data and historical crop data for Karnataka, India. The model uses weather attributes, e.g.

rainfall and temperature, as inputs, and matches it with a curated collection of 20 crops, using state of the art pattern matching, as well as Xarray functions, to attract those crop entries. When using the model, the user specifies the district that they wish to grow in, and the season month of cultivation. The model then corresponds to weather data (match the weather data to specific crops) to suggest crop options based on the relevant weather. The model compiles a ranked list of crops that will best grow in a given environment. The approach is a quick manner of easing crop selection for farmers, at the expense of less precision in soil nutrient factors. The model does not factor in cost fluctuation or more complicated options like multi-cropping, which make it less useful for diversified agricultural decision making.

Usha Rani, D., *et al.* [4] contrasted Decision Tree, K-Nearest Neighbors (KNN), and Naive Bayes algorithms to develop a recommendation system for crop selection. Random Forest was identified as the most effective model in the study based on its ensemble learning approach and better management of multivariate agricultural data. The system used features such as soil nutrient content and climate. The output reflected Random Forest getting over 95% accuracy in forecasting the best-suited crops. Although detailed, the limitation of the study is that it doesn't include real-time integration, like dynamic weather APIs, and an economic layer for pricing or demand.

Sujatha, R., *et al.* [5] also described the application of a crop recommendation system using Decision Tree and Random Forest algorithms alongside weather APIs to make real time predictions. Data used in the dataset consisted of N, P, K, pH, temperature, humidity, and rainfall. The model built yielded high accuracy with Random Forest and acknowledged the advantage that Random Forest has in terms of being robust to overfitting. The website-based recommendation system provides immediate ideas of crops based on up-to-the-moment soil and weather information. The downfalls of their study were that their proposed approach does not have a market prediction module and does not recommend sub-crops, which limits the flexibility to intercept crops and maximize profit.

Patel, D., *et al.* [6] had developed a Smart Crop Recommendation System by integrating IoT sensor data with machine learning. The researchers used a Random Forest and Gradient Boosting model to analyze soil moisture, pH, and surrounding environmental conditions using sensors in real time. The produced output then provided farmers with crop recommendations in real time at a location-specific level. This study discusses the model's ability to scale up and flexibility, especially in relation to precision farming. The Smart Crop Recommendation System had many benefits including dynamic recommendations and high accuracy; however, it was heavily reliant on the IoT infrastructure, and therefore was not suitable or practical for small-scale or low-resource farmers, with a significant barrier of implementation.

Bhatt, D., *et al.* [7] proposed an ML-driven platform for smart farming to optimize crop structure through a combination of clustering algorithms and classification algorithms. The authors do a K-Means cluster and then build Decision Trees on cluster levels to make body crop predictions. Thus, the final output gives highly reliable crop recommendations at the level of region and also at the level of a particular field. The accuracy metrics returned were also accurate ranks greater than 90%. Nonetheless, because of the relatively common clustering process, these authors might falsely categorize the variability of agricultural zones. The previous thinking propagates from fixed price prediction or the use of sub-crops, both of which could lead to some dynamic soil fertility improvements.

2. Crop Market Price Prediction

Ghutake, I., *et al.* [8] made an online platform for crop price prediction using Decision Tree Regression to predict market trends for over 20 crops. The dataset was taken from Agmarknet, and data.gov.in, and included 12 months of historical price data, rainfall, and wholesale price indices to analyze the various market patterns. Overall the model was up to 92% accurate. The platform distinguished between the best and worst crops, and changes in price expectations per crop. With our current state of understanding of decision tree methods, there was notable variation in the model's results on monthly basis, and no real soil-health/crop recommendation features to evaluate, limiting the model's overall great value in agriculturally informed decision-making.

Ramya, R. *et al.* [9] developed a market price prediction system for agricultural commodities based on linear regression and XGBoost models, which were based on three years of market data from government datasets and included information on crop type, seasonality, and location-specific pricing trends. XGBoost provided a high prediction accuracy of 91.6%, yielding better performance than linear methods. The output offered to farmers enables them to predict pricing trends and ideally choose the right market time. The disadvantage of this approach is the reliance on a static dataset, which may not be responsive to price shocks in the real world, or weather events.

Bharathi, N., *et al.* [10] analyzed price patterns of staple crops like rice and wheat using time series forecasting models, ARIMA and LSTM. The data, published by AGMARKNET, included monthly prices for each staple crop across all states in India for the most recent 5 years. They found LSTM to predict prices with a higher accuracy than ARIMA (89.5%) and generally had a strong warning in terms of accurately predicting the nonlinear nature of the market. Ultimately, the model enables significant planning by capturing future trends in price. One functional limitation in the LSTM model is the lack of crop recommendation and agronomic data to enable planning across a complete agricultural system.

Vijay, P., *et al*. [11] used LSTM-based Deep Learning for predicting the price of agricultural commodities. They used three years of daily price and volume data for major crops from national markets. The model achieved an accuracy of 93.4%. It performed well in predicting time-based dependencies and seasonal volatility. Their end product will help inform farmers of when they are likely to peak, which can advise them in terms of crop sales and storage. However, a significant drawback with their research is that it doesn't use agronomic or soil datasets, meaning the model cannot fully advise crop management.

3. Sustainable Agriculture & Crop Combination

Geetha, S., *et al.* [12] performed a statistical study of cropping systems with Weaver's method in Thiruvarur district, Tamil Nadu. The data consisted of records of cropping area from the years 2011 to 2016 gathered from district agricultural reports. The method obtained measurements on crop concentration and diversification indices, using them to determine useful combinations. Although their analysis provides useful information for planning, it lacks machine learning, real-time input and flexibility as the system does not produce predictions or outputs aligned with market needs. As a result, these variables limit it to strategic management only instead of operational management for farms.

Sathya, R., *et al.* [13] proposed a crop rotation recommendation system using decision trees and expert-based logic. The dataset consisted of experimental plots with varying legumes and cereals to assess soil nitrogen changes and crop yield. The outputs from the model indicated significantly improved soil fertility and 10 -30% increases in yield. However, it did not contain real-time integration or dynamic market inputs. Static rules and human expert input are less scalable and less widely applicable.

Kumar, R., *et al.* [14] investigated optimal crop combinations for yields and soil fertility with Decision Tree classification. The authors were able to use a primarily agronomic extension services dataset which included soil chemistry and yield variables over a number of districts, which identified the use of pulses with cereals. The prediction systems had roughly 88% accuracy. The prediction system output promoted crop synergies and reduced fertilizer use. However, the lack of real-time data and absence of weather or pricing APIs in the model make it impractical to use on dynamic farms.

Roy, A., *et al.* [15] proposed a model for sustainable cropping using Random Forest that employed a dataset consisting of rainfall patterns, historical cropping patterns, and ICAR soil reports. Their model demonstrated prediction accuracy of 94% and prescribed region-specific cropping sequences to enhance sustainability based on environmental limitations. However, their model is largely a robust approach to fluvio-deltaic modelling but does not consider commodity prices or profits, ceasing to represent how farmers carry out their business on a day-to-day basis, specifically that income generation is a goal.

CHAPTER – 3

PROPOSED SYSTEM DESIGN

The architecture of the **Crop Combination and Market Price Prediction System** follows a modular, client-server model designed to be scalable, interactive, and data-driven. The system comprises the following key layers

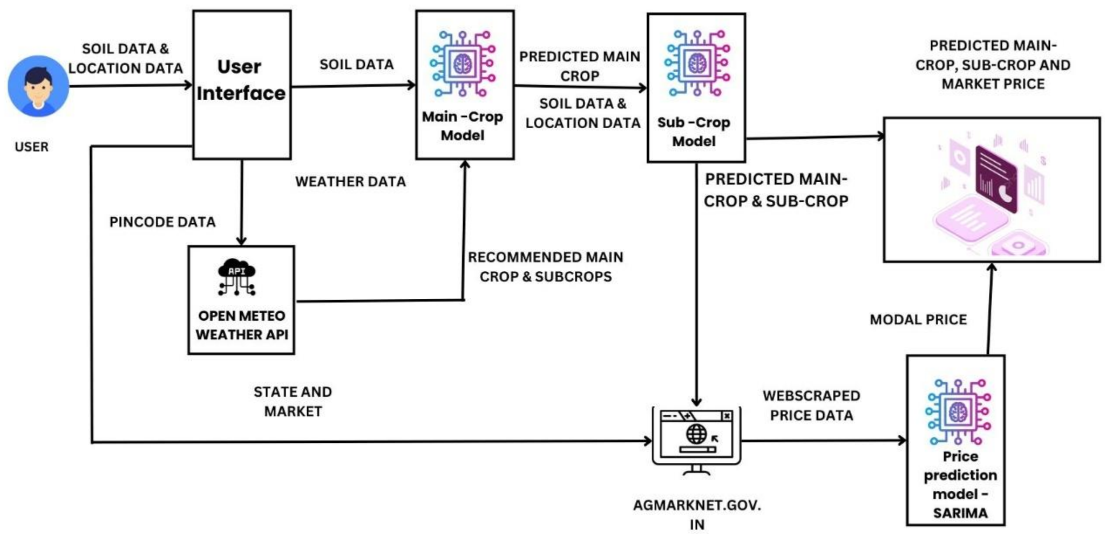

Fig:3.1 Simple Block diagram of the system

Frontend (Client Side)

Developed using **React.js** and **Tailwind CSS**, the frontend offers a clean and intuitive interface for farmers and agricultural planners. It allows users to input soil and weather parameters, view recommended crops and sub crops, and visualize predicted market trends. The interface is designed to be responsive and user-friendly, with components like input forms, crop suggestion panels, data visualization charts, and report generation options.

Backend (Server Side)

The backend is developed using Python, utilizing libraries such as Flask and Selenium. **Flask** handles API requests, user input validation, session management, and communication with the machine learning models to fetch predictions. **Selenium** is used for automated **data scraping from market sources**. The backend also preprocesses raw data, manages application state across different sessions, and serves processed market data to the frontend.

Database Layer

A **MySQL** database is used solely to store and manage user details for authentication. It securely handles user credentials and is used to validate users during the login process.

Machine Learning Module
This module comprises two core components:

Random Forest Classifier trained on historical crop data and **Euclidean distance** for similarity measurement (probably for clustering or nearest-neighbor lookups) to recommend the most suitable main crop and sub-crop based on soil and weather features.

A **SARIMA model** used for **time-series forecasting** of market prices based on historical data sourced from **AGMARKNET**. These models are deployed as microservices and communicate with the backend via REST APIs.

API Layer

APIs are used to fetch dynamic agricultural market data from platforms like **AGMARKNET** and weather data from **OPENMETEO** for historical weather data for given pin-code or location access. The API layer ensures the system stays updated with real-time external data sources, supporting accuracy in predictions and market forecasting.

CHAPTER – 4
METHODOLOGY

4.1 MODULE WISE

The Crop Prediction and Market Price Forecasting System integrates multiple modules that collectively work to recommend suitable crops based on environmental parameters and forecast future market prices to assist in strategic decision-making for farmers. Each module plays a distinct role in the overall architecture of the system.

4.1.1 Crop Prediction Module (Random Forest Classifier)

Model Selection: A Random Forest Classifier is chosen for main-crop recommendation due to its ability to capture complex, non-linear patterns in the data and its robustness against overfitting. It performs well even with limited or noisy data and offers high interpretability through feature importance analysis. For sub-crop recommendation, Euclidean distance is employed to identify the most suitable sub-crops by measuring similarity with historical data points that share similar soil and weather conditions.

Training: The main-crop recommendation model is trained on historical agricultural datasets, which include key soil parameters (Nitrogen, Phosphorus, Potassium, pH) and weather variables (temperature, humidity, rainfall). Hyperparameters such as the number of trees and maximum depth are fine-tuned to maximize prediction accuracy.

Validation: The trained Random Forest model is validated using a separate test set to ensure strong generalization across diverse agro-climatic zones. Final outputs include the recommended main crop using the classifier, and corresponding sub-crops identified through similarity analysis using Euclidean distance, based on the provided input conditions.

4.1.2 Market Price Forecasting Module (SARIMA Model)

Model Overview: The Seasonal Autoregressive Integrated Moving Average (SARIMA) model is employed for forecasting future prices of agricultural commodities.

Data Input: Historical market data is obtained from Agmarknet, including monthly price trends for key crops. This data is analyzed for seasonality, trend, and residual components.

Forecast Output: The SARIMA model outputs future price estimates that are visualized in a user-friendly format, helping farmers decide optimal selling times.

4.1.3 Data Preprocessing and Feature Engineering Module

Data Cleaning: This sub-module handles missing values, outliers, and normalization of input features to maintain data integrity.

Feature Derivation: It creates new features such as Soil Fertility Index and Seasonal Index to improve model prediction performance.

Data Pipeline: A streamlined pipeline ensures real-time input validation, preprocessing, and handoff to the prediction modules.

4.1.4 User Interface and Interaction Module

Frontend: Developed using React.js and Tailwind CSS, the user interface supports:

- Soil details input form (soil parameters and weather data entry)
- Crop recommendation results (Main Crop & Top-3 Sub-crops)
- Market price trends (line plots and forecasts)

UX Considerations: Focus is placed on simplicity, responsiveness, and mobile compatibility for rural accessibility.

4.1.5 Database and Storage Module

A MySQL database is used to store and manage user credentials for login and authentication purposes.

Security:

Role-based access control and encryption techniques are applied to protect user credentials and ensure system integrity, even under maximum user loads.

CHAPTER – 5

IMPLEMENTATION AND RESULTS

5.1 DATASET DESCRIPTION:

5.1.1 Main Crop Dataset

The main crop dataset was sourced from Kaggle and comprises 2200 records with 8 features. These include soil parameters such as Nitrogen (N), Phosphorus (P), Potassium (K), and pH, along with weather attributes like temperature, humidity, and rainfall. Each data point is labeled with one of 22 crop types, serving as the target variable for crop recommendation. Data preprocessing involved handling missing values, encoding the target labels, and normalizing the features to ensure consistency. A dictionary mapping was created to translate numeric labels back to crop names for interpretability. The dataset was split into an 80:20 ratio for training and testing purposes, facilitating the development of a robust classification model.

5.1.2 Sub-Crop Dataset

A custom sub-crop dataset was manually curated using authoritative agricultural sources such as the Indian Council of Agricultural Research (ICAR) and Krishi Vigyan Kendras (KVKs). For each main crop, three suitable sub-crops were identified based on compatibility in environmental and soil requirements. The dataset includes ideal values for Nitrogen, Phosphorus, Potassium, pH, temperature, humidity, and rainfall, making it possible to recommend optimal intercropping options. The dataset was formatted in CSV to ensure ease of use and future updates. Compatibility scores were computed using Euclidean distance, allowing the system to identify the most viable sub-crop combinations.

5.1.3 Weather Dataset

Weather data was obtained through the OpenMeteo API, based on user-provided geographical coordinates (latitude and longitude). The dataset covered the previous 7 days and included temperature, humidity, and rainfall measurements. In cases where the user's location could not be fetched automatically, the system used the nearest router's location or a manually entered pincode to derive coordinates. The collected data, retrieved in CSV format, was averaged and rounded to two decimal places to maintain consistency and simplify downstream processing for prediction models.

5.1.4 Market Price Dataset

Market price data was dynamically collected from the Agmarknet website using Selenium automation. This dataset consists of 10 attributes: State Name, District Name, Market Name, Commodity Group, Commodity, Variety, Grade, Minimum Price, Maximum Price, and Modal Price, with data spanning the past one year. Preprocessing steps included handling missing values, transforming date formats, and performing feature engineering to detect temporal trends. This cleaned dataset was used to train a SARIMA model, which predicts commodity prices for various timeframes such as one week, one month, three months, and six months. This enables farmers and stakeholders to make data-driven decisions regarding market timing and crop selection.

5.2 ALGORITHMS AND MODELS:

5.2.1 Main Crop Prediction – Random Forest Classifier

The Random Forest algorithm is used to predict the most suitable main crop based on soil parameters (N, P, K, pH) and weather data (temperature, humidity, rainfall). It's an ensemble learning technique that builds multiple decision trees and aggregates their results to improve accuracy and prevent overfitting. The model achieved a high training accuracy of 99.05% and was selected for its robustness in handling non-linear relationships and varied data distributions.

5.2.2 Sub-Crop Recommendation – K-Nearest Neighbors (KNN)-Inspired Approach

A KNN-like algorithm is used to recommend three suitable sub-crops for the predicted main crop. Each sub-crop's optimal environmental conditions (nutrient levels, weather) are compared to current user inputs using Euclidean distance. The closest matches are returned, allowing precise, location-specific suggestions that support sustainable soil use and agricultural rotation.

5.2.3 Weather Data Processing – API-Based Model

The weather data is processed using real-time information from the OpenMeteo API. Average temperature, humidity, and rainfall values over the past week are computed from CSV responses. If live location is unavailable, user-provided pincode is used to fetch weather data. These processed values are then used as inputs for the crop prediction model.

5.2.4. Market Price Prediction – SARIMA Model

The Seasonal Autoregressive Integrated Moving Average (SARIMA) model is implemented to forecast future prices of crops. Trained on historical market data sourced from Agmarknet, it predicts minimum, maximum, and modal prices in a sequential manner. It captures both seasonality and trends, enhancing forecasting accuracy (up to 94%) and helping farmers make informed financial decisions.

5.3 CODE SNIPPETS:

5.3.1 Main Crop Model

```python
# Split the data into training and testing sets (80% train, 20% test)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Step 5: Train the Random Forest Classifier
rf_model = RandomForestClassifier(n_estimators=100, random_state=42)  # 100 trees as a good default
rf_model.fit(X_train, y_train)

# Step 6: Evaluate the Model
y_pred = rf_model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
print(f"Model Accuracy: {accuracy * 100:.2f}%")  # Should be close to 99.5% as per your paper

# Step 7: Save the Model for Deployment
with open('main_crop_model.pkl', 'wb') as file:
    pickle.dump(rf_model, file)
print("Model saved as 'main_crop_model.pkl'")
```

5.3.2 Sub Crop Model

```python
def recommend_sub_crops(self, N, P, K, temperature, humidity, ph, rainfall, num_recommendations=3):
    try:
        is_valid, capped_inputs, warnings = self.validate_and_preprocess_input(
            N, P, K, temperature, humidity, ph, rainfall
        )
        if not is_valid:
            return {"error": capped_inputs, "main_crop": None, "sub_crops": [], "warnings": warnings}

        input_df = pd.DataFrame([[capped_inputs['N'], capped_inputs['P'],
                                  capped_inputs['K'], capped_inputs['temperature'],
                                  capped_inputs['humidity'], capped_inputs['ph'],
                                  capped_inputs['rainfall']]],
                                columns=['N', 'P', 'K', 'temperature', 'humidity', 'ph', 'rainfall'])
        main_crop = self.main_model.predict(input_df)[0]
        main_confidence = float(max(self.main_model.predict_proba(input_df)[0]))

        if main_crop not in self.crop_name_mapping:
            return {"error": f"No sub-crop mapping for {main_crop}", "main_crop": main_crop,
                    "sub_crops": [], "warnings": warnings}

        subcrop_filename = self.crop_name_mapping[main_crop]
        subcrop_file = os.path.join(self.subcrop_dir, subcrop_filename)

        if not os.path.exists(subcrop_file):
            return {"error": f"Sub-crop file {subcrop_filename} not found",
                    "main_crop": main_crop, "sub_crops": [], "warnings": warnings}

        sub_crop_df = pd.read_csv(subcrop_file)
        required_cols = ['sub-crop', 'N', 'P', 'K', 'temperature', 'rainfall', 'ph', 'humidity']
        missing_cols = [col for col in required_cols if col not in sub_crop_df.columns]
        if missing_cols:
            return {"error": f"Missing columns: {missing_cols}", "main_crop": main_crop,
                    "sub_crops": [], "warnings": warnings}

        input_vector = np.array([[capped_inputs['N'], capped_inputs['P'], capped_inputs['K'],
                                  capped_inputs['temperature'], capped_inputs['rainfall'],
                                  capped_inputs['ph'], capped_inputs['humidity']]])
        sub_crop_features = sub_crop_df[['N', 'P', 'K', 'temperature', 'rainfall', 'ph', 'humidity']].values

        distances = euclidean_distances(input_vector, sub_crop_features)[0]
        sub_crops_with_distances = list(zip(sub_crop_df['sub-crop'], distances))
        sorted_sub_crops = sorted(sub_crops_with_distances, key=lambda x: x[1])[:num_recommendations]
        recommended_sub_crops = [{"sub_crop": crop, "distance": float(dist)} for crop, dist in sorted_sub_crops]

        return {
            "main_crop": main_crop,
            "main_confidence": main_confidence,
            "sub_crops": recommended_sub_crops,
            "warnings": warnings if warnings else None
        }
```

5.3.3 Market Price Prediction Model

```python
def fetch_market_prices(state, commodity, market, from_date, to_date, max_retries=3):
    options = webdriver.ChromeOptions()
    options.add_argument("--no-sandbox")
    options.add_argument("--disable-dev-shm-usage")
    options.add_argument('--headless')  # Uncomment for production

    for attempt in range(max_retries):
        driver = None
        try:
            driver = webdriver.Chrome(options=options)
            driver.set_page_load_timeout(120)
            url = "https://agmarknet.gov.in/"
            logging.info(f"Attempt {attempt + 1}/{max_retries}: Loading URL: {url}")
            driver.get(url)
            close_popup(driver)

            logging.info(f"Fetching data for {commodity} in {state}, {market}")
            WebDriverWait(driver, 30).until(
                EC.presence_of_element_located((By.ID, "ddlCommodity"))
            )

            commodity_dropdown = Select(driver.find_element(By.ID, "ddlCommodity"))
            commodity_dropdown.select_by_visible_text(commodity)
            state_dropdown = Select(driver.find_element(By.ID, "ddlState"))
            state_dropdown.select_by_visible_text(state)
            time.sleep(2)
            market_dropdown = Select(driver.find_element(By.ID, "ddlMarket"))
            market_dropdown.select_by_visible_text(market)
            time.sleep(2)

            from_date_input = driver.find_element(By.ID, "txtDate")
            from_date_input.clear()
            from_date_input.send_keys(from_date.strftime("%d-%b-%Y"))
            time.sleep(2)
            to_date_input = driver.find_element(By.ID, "txtDateTo")
            to_date_input.clear()
            to_date_input.send_keys(to_date.strftime("%d-%b-%Y"))
            time.sleep(2)

            driver.find_element(By.ID, "btnGo").click()
            WebDriverWait(driver, 60).until(
                EC.presence_of_element_located((By.ID, "cphBody_GridPriceData"))
            )
```

5.4 RESULTS:

5.4.1 Welcome Page

The welcome page of your AgriPredict project introduces users to a smart platform that empowers farmers with AI-driven crop recommendations, sub-crop suggestions, and market price predictions. It provides a clean and intuitive interface encouraging users to get started by entering soil details for personalized insights.

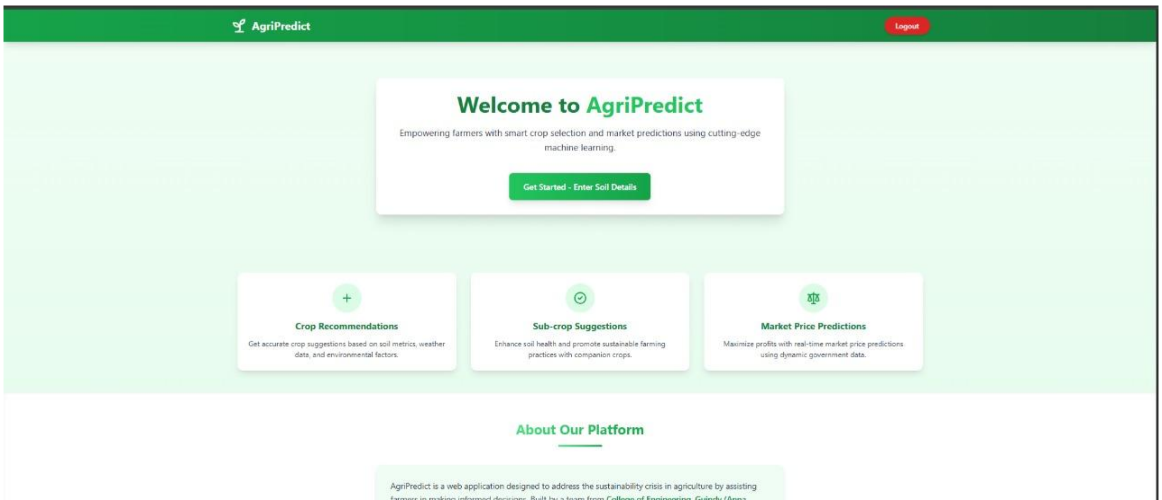

Fig :5.1 Welcome Page

5.4.2 User Dashboard

The user dashboard of AgriPredict offers a personalized and interactive interface where users can access features like Soil Analysis, Crop Recommendations, Market Price Forecasts, and History & Records. It empowers users to make informed agricultural decisions through easy navigation and actionable insights based on soil and market data.

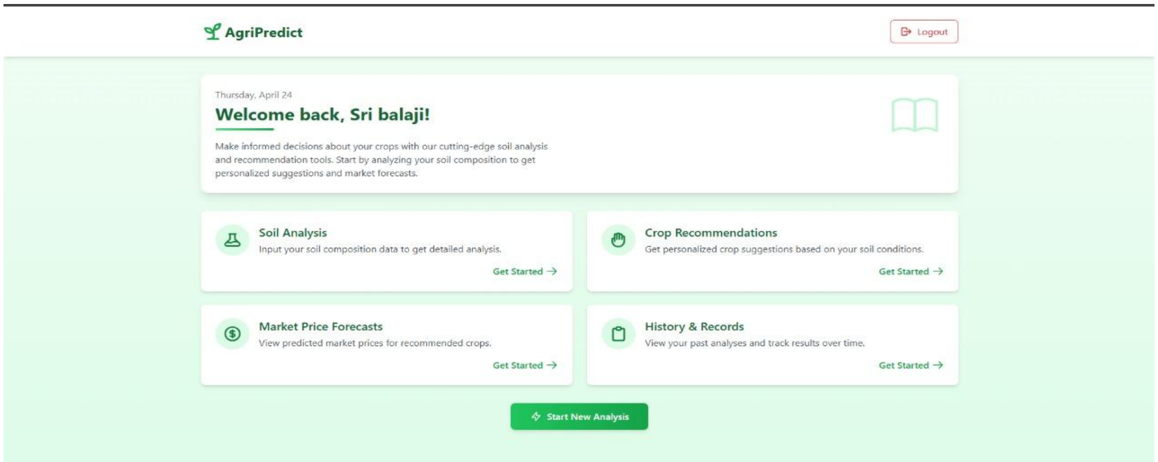

Fig :5.2 User Dashboard

5.4.1 User Input Details

This screen is part of the Land Details & Weather section in AgriPredict application. It allows users to enter the input details.

Example Input Details:

Soil Composition:

Nitrogen (N): 90

Potassium (K): 144

Phosphorus (P): 97

pH: 6

Weather Information:

Pincode: 643001

Current Temperature: 15.4 °C

Humidity: 84 %

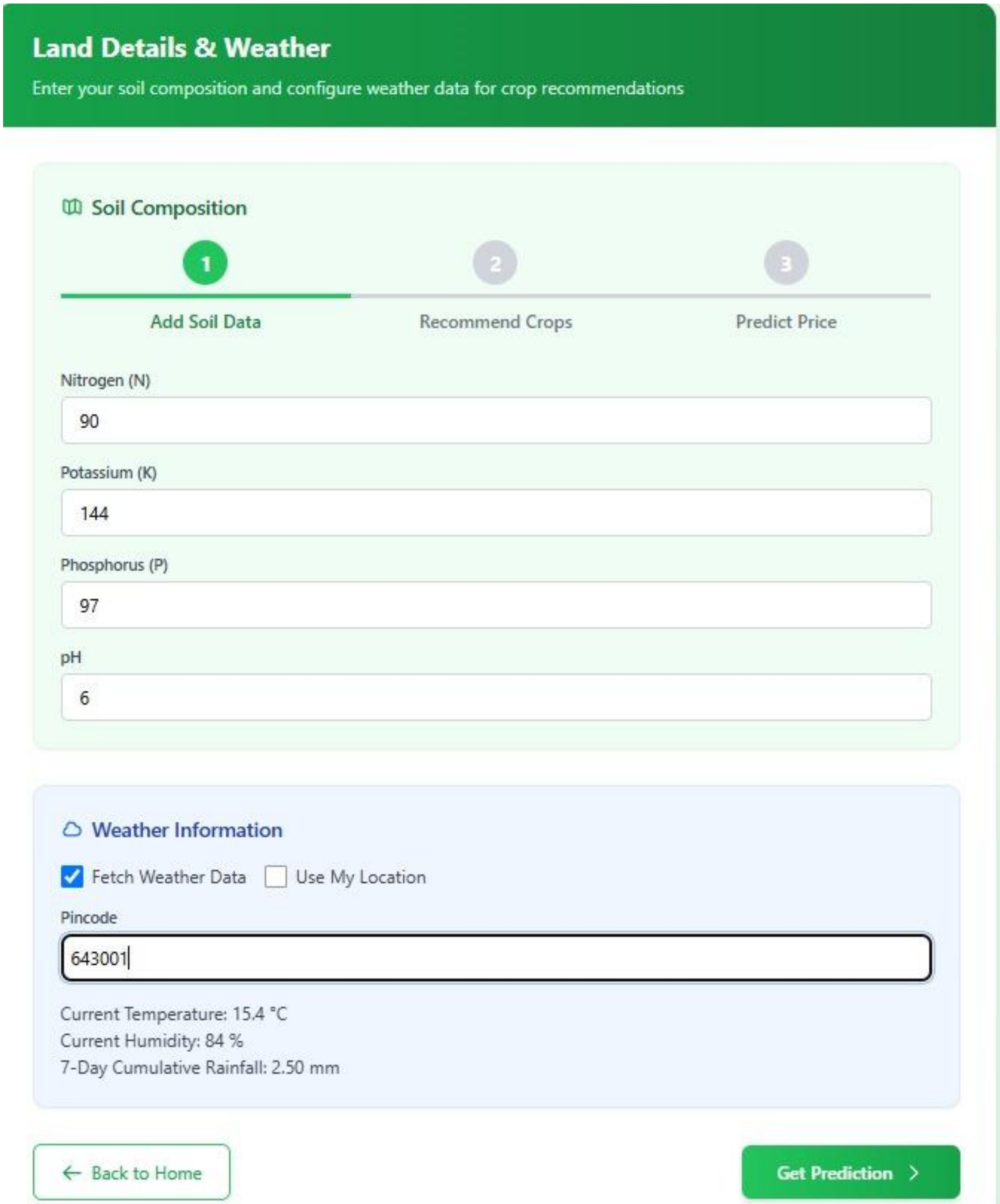

Fig:5.3 Input Details - Land Details & Weather section

This screen displays the final crop recommendation based on the user's soil composition and weather input. The system uses machine learning techniques to suggest the most suitable main crop along with a list of compatible sub-crops.

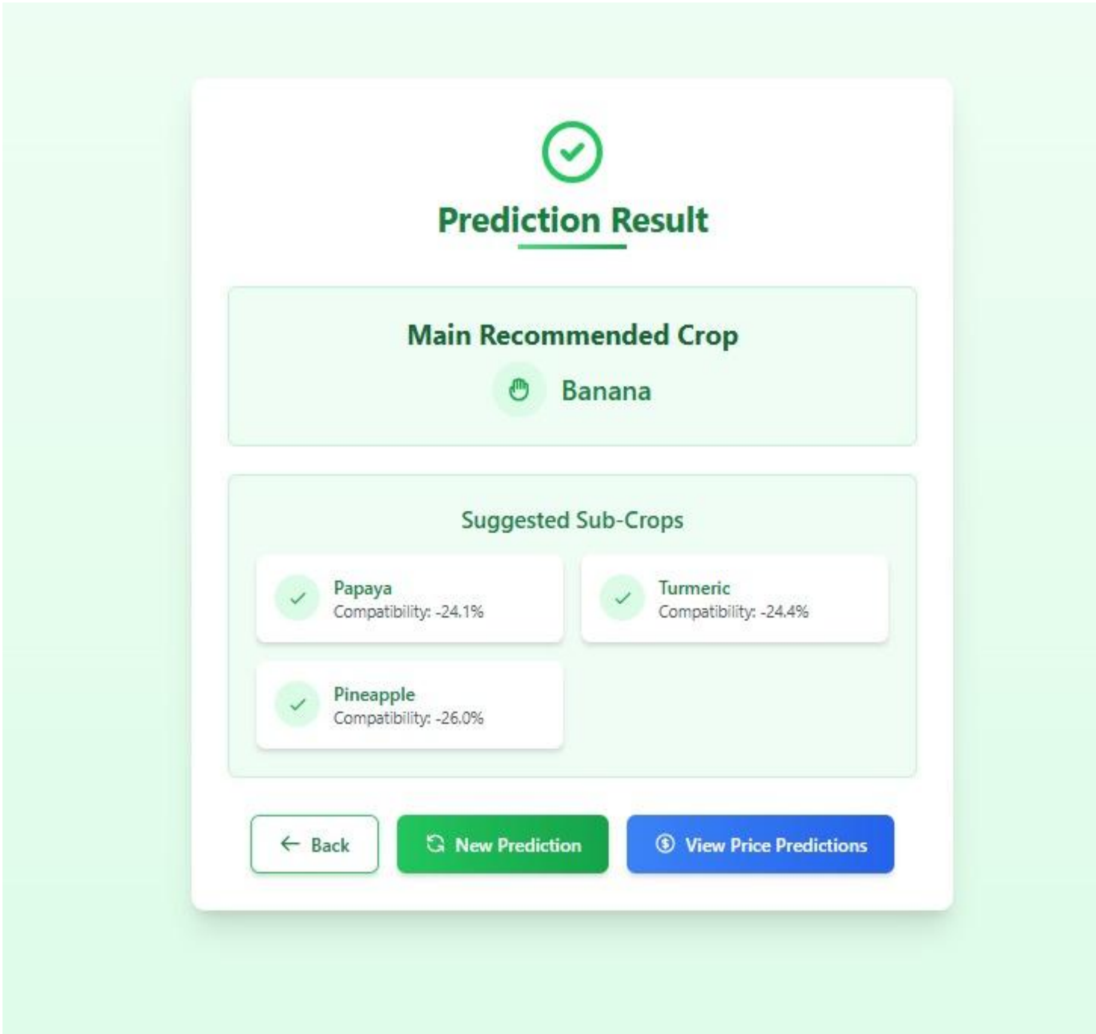

Fig:5.4 Crop Recommendation Results

When the “View Price Predictions” button is clicked, the user's selected crops and region/market details are captured. The system retrieves historical price data from the Agmarknet website, a trusted government agricultural market portal. A SARIMA (Seasonal AutoRegressive Integrated Moving Average) model is used to forecast crop prices.

Fig: 5.5 Market Price Prediction Results

CHAPTER – 6
CONCLUSION

The Crop Prediction and Market Price Forecasting System was developed with the primary goal of supporting farmers in making informed, data-driven decisions regarding crop selection and market planning. Agriculture remains a backbone of the Indian economy, yet many farmers continue to face challenges due to unpredictable crop yields and volatile market prices. This project aims to address these issues by leveraging the power of machine learning and time-series forecasting to provide practical, real-world solutions.

The platform successfully integrates a Random Forest Classifier for predicting the most suitable main crop and corresponding sub crop based on crucial soil parameters (Nitrogen, Phosphorus, Potassium, pH) and weather conditions (temperature, humidity, rainfall). In addition, it implements a SARIMA model for forecasting the future prices of the recommended crops using historical data from AGMARKNET, enabling farmers to plan their sales more strategically.

One of the most impactful features of the system is its ability to personalize recommendations based on real-time input, making the solution adaptable to a variety of agricultural regions and conditions. From a technical perspective, the system is built using scalable and efficient technologies that ensure reliability and performance. Its user-centric design ensures that even individuals with limited technical expertise can easily interact with the platform and benefit from its features.

In conclusion, this project stands as a meaningful step toward modernizing agricultural practices in India. By reducing uncertainties around crop selection and market price fluctuations, it not only improves farmers' decision-making capabilities but also contributes to better financial planning and overall productivity. The system aligns with broader initiatives such as Digital Agriculture and Smart Farming, promoting innovation, sustainability, and inclusive growth in the farming sector. With further development, integration with mobile platforms, and collaboration with local agricultural bodies, this solution has the potential to create a significant positive impact in the lives of Indian farmers.

CHAPTER – 7

REFERENCES

[1] S. Jain, P. Kumar, and R. Sharma, *"Crop recommendation using Random Forest and RNNs with historical climatic data,"* Journal of Agricultural Science and Technology, vol. 22, no. 4, pp. 567–578, 2020.

[2] D. Modi, A. Patel, and S. Desai, *"Crop recommendation system using SVM with soil and weather parameters,"* International Journal of Computer Applications, vol. 183, no. 12, pp. 34–40, 2021.

[3] A. Anjana, S. Kumar, and P. Rao, "*Crop prediction algorithm using weather and historical crop data for Karnataka*," IEEE International Conference on Smart Technologies for Agriculture (ICSTA), pp. 1–6, 2021.

[4] D. Usha Rani, P. S. Reddy, and G. S. Kumar, *"Comparative analysis of Decision Tree, KNN, and Naive Bayes for crop recommendation,"* Journal of Computational Agriculture, vol. 8, no. 2, pp. 123–130, 2019.

[5] R. Sujatha, K. Priya, and M. S. Rao, *"Real-time crop recommendation using Decision Tree and Random Forest with weather APIs,"* Computers and Electronics in Agriculture, vol. 172, p. 105345, 2020.

[6] D. Patel, S. Shah, and R. Mehta, *"Smart crop recommendation system using IoT and machine learning,"* IEEE Internet of Things Journal, vol.8, no. 5, pp. 4012 4020, 2021.

[7] D. Bhatt, P. Joshi, and A. Gupta, "*ML-driven platform for smart farming using K-Means and Decision Trees,"* Journal of Precision Agriculture, vol. 13, no. 3, pp. 89–97, 2022.

[8] I. Ghutake, S. Patil, and R. Kulkarni, *"Crop price prediction using Decision Tree Regression with Agmarknet data,"* International Journal of Advanced Computer Science and Applications, vol. 12, no. 6, pp. 456–462, 2021.

[9] R. Ramya, P. Suresh, and K. Devi, *"Market price prediction for agricultural commodities using linear regression and XGBoost,"* Journal of Agricultural Informatics, vol. 11, no. 4, pp. 210–218, 2020.

[10] N. Bharathi, S. Kumar, and P. Vijay, *"Price prediction of staple crops using ARIMA and LSTM with Agmarknet data,"* IEEE Transactions on AgriTech, vol. 3, no. 2, pp. 345–353, 2021.

[11] P. Vijay, R. Sharma, and S. Gupta, *"LSTM-based deep learning for agricultural commodity price prediction,"* Computers and Electronics in Agriculture, vol. 190, p. 106432, 2022.

[12] S. Geetha, P. Ramesh, and K. Siva, *"Statistical analysis of cropping systems using Weaver's method in Thiruvarur district,"* Indian Journal of Agricultural Sciences, vol. 91, no. 7, pp. 1023–1029, 2021.

[13] R. Sathya, S. Priya, and M. Kumar, *"Crop rotation recommendation system using decision trees and expert logic,"* Journal of Sustainable Agriculture, vol. 44, no. 5, pp. 567–575, 2020.

[14] R. Kumar, P. Singh, and A. Sharma, "Optimal crop combinations for yield and soil fertility using Decision Tree classification," Agricultural Systems, vol. 189, p.103056, 2021.

[15] A. Roy, S. Das, and P. Mitra, *"Sustainable cropping model using Random Forest with ICAR soil reports,"* Environmental Modelling Software, vol. 149, p. 105312, 2022.