2024 IEEE 9th International Conference for Convergence in Technology (I2CT)
Pune, India. Apr 5-7, 2024

Crop Combination And Market Prediction using ML methods

Sree Nidhi M
Computer Science
PES University
Bangalore, India
sreenidhi8765@gmail.com

Ashwathy S Kumar
Computer Science
PES University
Bangalore, India
skumarashwathy23@gmail.com

Tarun S
Computer Science
PES University
Bangalore, India
legend432002@gmail.com

Dinesh P Kumar
Computer Science
PES University
Bangalore, India
chowdaryd365@gmail.com

Dr. Prajwala TR
Associate Professor
Computer Science Department
PES University
Bangalore, India
prajwalatr@pes.edu

2024 IEEE 9th International Conference for Convergence in Technology (I2CT) | 979-8-3503-9447-4/24/$31.00 ©2024 IEEE | DOI: 10.1109/I2CT61223.2024.10543457

***Abstract*—Agriculture's sustainability crisis, marked by soil degradation and unbalanced ecosystems, calls for innovative solutions. Our web application addresses this by aiding farmers in crop selection and market prediction, rooted in the principles of healthier cropping methods. Crop rotation, a time-honored technique dating back to ancient civilizations, has been scientifically proven to combat soil nutrient depletion, pest infestations, and erosion, crucial for long-term land productivity. The application's recommendation of leguminous sub-crops, like soybeans, plays a pivotal role in enriching soils with nitrogen, a key factor in increasing yields of subsequent crops without the excessive use of synthetic fertilizers. Studies, including one by Washington State University, have shown that crop rotation can significantly increase harvests, with some cases reporting up to a 58% yield increase. Integrating these agricultural best practices with advanced predictive models for market prices, our application empowers farmers with knowledge to enhance soil health, boost crop yields, and make informed market decisions, thereby contributing substantially to sustainable farming and economic stability.**

***Index Terms*—Sustainable agriculture, Soil degradation, Ecosystem balance, Crop selection, Market prediction, Crop rotation, Soil nutrient management, Pest infestation control, Erosion prevention, Land productivity, Leguminous crops, Nitrogen fixation, Synthetic fertilizers reduction, Machine learning models, Predictive analytics, Market price forecasting, Agricultural best practices, Environmental sustainability, Economic stability, Agronomy, Data science applications, Random Forest algorithm, Soil health improvement, Precision farming, Sustainable farming practices.**

I. Introduction

In an era where technological advancement is at its zenith, agriculture, the bedrock of human civilization, faces a paradox. On one hand, there's an ever- growing demand for food driven by the increasing global population; on the other, modern agricultural practices are encountering unprecedented challenges.In the world of farming, where challenges like changing weather, soil problems, and uncertain market demands make it tough to grow crops, using machine learning to predict the best crop combinations offers hope. These challenges range from soil degradation and pest infestations to fluctuating market prices, all of which significantly impact crop yield and the economic stability of farmers. Our project presents an innovative solution to these issues: a web application meticulously designed to assist farmers in making informed decisions about crop selection and market dynamics.

The inception of this project was fueled by the critical need to integrate sustainable farming methods with modern technology. Historically, agriculture has been the lifeline of societies, but conventional practices have often led to detrimental effects on soil health and the environment. Recognizing the importance of soil as a fundamental resource for agriculture, our application is a stride to- wards reversing soil degradation. It leverages the principles of healthier cropping methods, particularly crop rotation, which has been an effective strategy since ancient times. This practice, backed by scientific research, has shown significant benefits in maintaining soil fertility, controlling pests, and preventing erosion, thereby ensuring long-term productivity and environmental balance.

Our application stands at the intersection of agronomy and data science. It utilizes a Flask-based framework, complementing Python's analytical prowess with HTML and CSS for a user-friendly interface. The core of the application is a machine learning model, specifically a Random Forest algorithm, adept at analyzing complex agricultural data. This model provides recommendations for a primary crop based on the farmer's specific soil and environmental conditions. However, our approach goes a step further by suggesting sub-crops that complement the main crop, thereby enriching the soil naturally, reducing the need for synthetic fertilizers, and promoting biodiversity.

The selection of sub-crops is not arbitrary but is deeply rooted in the principles of crop rotation and inter-cropping. For instance, the inclusion of leguminous crops, known for their nitrogen-fixing capabilities, plays a crucial role in replenishing soil nutrients. This method of natural soil enrichment not only boosts the yield of subsequent crops but also aligns with sustainable farming practices, a necessity in today's ecologically conscious world.

Another pivotal feature of our application is the market price prediction tool. In the volatile realm of agricultural markets, having foresight into crop prices can be a game-changer for farmers. Our application addresses this by providing real-time price forecasts, utilizing dynamically updated datasets. This feature empowers farmers with critical insights, enabling them to plan their crop sales strategically, potentially maximizing their profits and mitigating economic risks. In summary, this paper will explore the intricate details of our application's development, from its conceptualization to its practical implementation. We will discuss the technological architecture, the machine learning models employed, and the user experience design. Moreover, we will delve into the potential impact of our application, both from an agronomic and an economic perspective. By intertwining historical agricultural practices with cutting-edge technology, our application is not just a tool but a testament to the synergy between tradition and innovation, aimed at empowering farmers towards sustainable and profitable farming.

II. RELATED WORK

In Paper [1], Sonal Jain and colleague propose a practical method for crop selection that considers both yearly and seasonal aspects. They use a Random Forest model trained on Telangana's agro-climatic data, combining it with Recurrent Neural Networks for weather forecasts and soil details to recommend the best-suited crops.The authors provide valuable insights for farmers, such as irrigation needs and ideal sowing times based on predicted weather. Despite its usefulness, the paper overlooks economic factors influencing crop choice and lacks consideration of crucial soil values (N, P, K), impacting crop suitability.

In Paper [2], Ishita Ghutake and team introduce an online Crop Price Prediction Website for farmers. This platform facilitates personalized crop choices based on financial conditions and feasibility, presenting the performance of numerous widely grown crops. It showcases the best and worst performers, highlighting their percentage increases or decreases over the next 12 months. Utilizing a structured dataset from government website[7], the website incorporates rainfall and wholesale price information for 20+ crops, employing decision tree regression in supervised learning. Despite its commendable 92% accuracy, variations month-to-month stand as a notable point. In their analysis discussed in Paper [3], S. Geetha and R. Maniyosai discuss about the cropping patterns, concentration, and diversification methods within Thiruvarur district, Tamilnadu. They specifically focus on how different crops are cultivated together in this region. Employing Weaver's method, they determine the most effective combinations of crops by calculating the minimum deviation. . The study, centered on Thiruvarur, highlights the potential of changing cropping approaches to enhance agricultural productivity. In the study

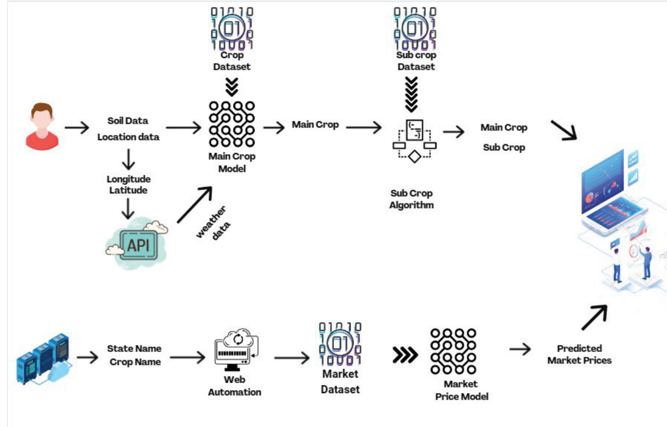

Fig. 1. System Architecture

This is the architecture diagram of our project which gives an overview of the high-level design depicting the interaction between the components of the system.

by Anjana, Aishwarya Kedlaya and others paper[4], a crop prediction system is proposed for Karnataka state. The dataset includes soil-specific and weather-related attributes gathered from IMD Pune[9] , with a manually curated crop dataset containing information on 20 major crops. The method relies on pattern matching and Xarray functions to analyze datasets. Farmers in Karnataka can input their district and desired cultivation month, and the system fetches and compares weather data with the crop dataset, displaying suitable crops.

In the study by Daneshwari Modi and other in paper [5] a dataset encompassing soil metrics such as N (nitrogen), P (phosphorus), K (potassium), pH, humidity, rainfall, and temperature is utilized for multi-class classification. The methodology, based on Support Vector Machines (SVM), factors in these soil parameters and environmental conditions to accurately recommend the most productive crops for specific soil characteristics. This approach assists farmers in making profitable crop choices, minimizing the risk of inaccuracies and boosting overall crop productivity. The proposed algorithm demonstrates an accuracy of 97%.

III. PROPOSED SYSTEM

The designed system includes different steps such as Data collection, Data visualization, Data preprocessing, classification and evaluation. Our system integrates four models—crop, weather API, sub-crop, and market. From user defined location,the latitude and longitude are used for weather API calls, generating parameters along with soil data to predict the main crop. Sub-crops are suggested based on the main crop, and market prices are predicted from government datasets, letting farmers to make informed crop choices as shown in figure 1.

A. Data Collection

Main crop:The dataset[8] contains 2200 rows and 8 columns, encompassing soil metrics including N, P, K, humidity, rainfall, pH, and temperature. These parameters are associated with various types of soil, each linked to a specific set of crops, denoted as labels. The dataset delineates the relationship between soil characteristics and the 'n' number of crops dependent on diverse soil parameters. Subcrop:Manually created sub-crop data based on insights gathered from various research sources. Data sources include reputable agricultural organizations such as the Indian Council of Agricultural Research (ICAR) [10]and Krishi Vigyan Kendras (KVKs)[11]. Market data: sourced in real-time using the Selenium tool from [6], provides an up-to-date market data about the crop. It contains 10 columns such as State Name, District Name, Market Name, Group, Commodity, Variety, Grade, Min Price (Rs/Quintal), Max Price (Rs/Quintal), Modal Price (Rs/Quintal), and Price Date. Retrospectively collected for a year until the present extraction date.

B. Data Preprocessing

Main Crop Prediction: In preparation for the crop recommendation model, we loaded the Kaggle dataset containing information about 22 crops. Essential steps included thorough exploration of the dataset, meticulous selection of relevant features such as nitrogen, phosphorous, potassium, temperature, humidity, pH, and rainfall, defining the target variable as the crop label, and appropriately splitting the data into training and testing sets. Employing a Random Forest Classifier on the selected features yielded an impressive accuracy of 99.54%. Additionally, we created a crop dictionary to map numeric labels to crop names, enhancing interpretability.

Sub-Crop Recommendation System: The proactive creation of datasets through extensive research ensures the availability of relevant and accurate information. We associated each distinct main-crop in the maincrop dataset with three suitable sub-crops. The efficient loading of CSV data promotes modularity and facilitates ease of dataset updates. Consequently, we established another dataset providing details on optimal nitrogen, phosphorous, potassium, temperature, humidity, pH, and rainfall values for the identified sub-crops.

Crop Market Price Prediction: Successful dynamic data retrieval from the Agmarknet website ensures real-time information, allowing our model to stay updated with the latest market changes. This approach contrasts with static datasets that may become pattern-bound and fail to capture real-time market fluctuations. Thoughtful data cleaning and feature engineering contribute to enriching the dataset with relevant temporal information, enhancing the accuracy of market price predictions.

C. Feature Engineering

Main Crop Prediction: The main crop prediction module lays a robust foundation by incorporating key features, namely Nitrogen (N), Phosphorous (P), Potassium (K), temperature, and humidity. These selected features form a critical framework for accurate crop prediction, enabling the Random Forest Classifier to discern intricate patterns and achieve a remarkable accuracy of 99.54%. The creation of a comprehensive crop dictionary, mapping numeric labels to crop names, further enhances the interpretability of the model outcomes.

Sub-Crop Recommendation System: Within the sub-crop recommendation system, a meticulous approach is taken, involving effective distance calculation and sorting mechanisms. This results in a well-defined hierarchy of recommended sub-crops tailored to the user's specific soil conditions. By leveraging this approach, the system provides farmers with precise and personalized suggestions for sub-crops that complement the main crop, fostering natural soil enrichment and sustainable farming practices.

Selection of Key Weather Features: Identification and selection of essential weather features, such as rainfall, humidity, and temperature, play a pivotal role in optimizing the model input for the crop recommendation system. The chosen weather features ensure that the model gains meaningful insights into the dynamic weather conditions that significantly influence crop growth. This meticulous selection process enhances the accuracy of the system's recommendations, aligning them closely with real-world agricultural scenarios.

Crop Market Price Prediction: In the crop market price prediction module, the inclusion of key features, including Minimum Price (Rs/Quintal), Maximum Price (Rs/Quintal), Modal Price (Rs/Quintal), and Price Date, is instrumental for achieving accurate market price predictions. This comprehensive set of features, combined with dynamic data retrieval from the Agmarknet website, ensures that the system is always equipped with the latest market information. Through thoughtful data cleaning and feature engineering, the dataset is enriched with relevant temporal information, contributing to the precision and reliability of market price forecasts.

D. Models And Implementation

- **Weather**: We collect weather data for a specified location (latitude, longitude) using the Visual Crossing Web Services API[10] over the past 30 days. Temperature, humidity, and rainfall values are extracted from the CSV response. Upon data availability, we calculate the average temperature, humidity, and monthly rainfall, presenting results in a two-decimal-place-rounded dictionary.Initially, the JavaScript code attempts to retrieve the user's current location coordinates.If successful, these coordinates are used for weather data retrieval.If unsuccessful, the code resorts to obtaining the nearest router location as a fallback option,If the user is dissatisfied with this location, they have the option to enter the pincode of their farming land.The pincode is then used to derive latitude and longitude values for the weather API call, ensuring accurate data retrieval. Access to weather data through the API requires an API key. This key facilitates access to the necessary weather data, allowing the generation of a URL for retrieving the last 30 days of information. With the API key and location coordinates, the weather data is retrieved from the specified API endpoint.Upon retrieval, the data is processed, initializing variables for cumulative values,average temperature, humidity, and rainfall are calculated based on the collected data.The calculated weather parameters (average temperature, humidity, and rainfall) are then fed into a crop model. The weather API is essential for accurate information, especially in agriculture, where data-driven decisions are key for smarter use of weather information and better outcomes.

- **Main crop prediction**: The selected predictive model for the main crop classification is the Random Forest algorithm, a supervised learning method renowned for its efficacy in handling complex datasets. This ensemble learning technique constructs numerous decision trees during training, capturing intricate relationships within the data. Three key attributes contribute to the superiority of Random Forest in our project:
  Random Subset Training: The model is trained on a random subset of the data, enhancing its adaptability and generalization to unseen data. Robust to Overfitting: Random Forest demonstrates resilience against overfitting, ensuring a reliable performance by avoiding excessive adaptation to the training data. Voting Mechanism: During predictions, each tree in the Random Forest contributes through a voting mechanism, either for class (classification problems) or regression output (regression problems), consolidating diverse perspectives for more accurate predictions. The selection of Random Forest proved advantageous, achieving an exceptional accuracy of 99.55%. Its robustness against overfitting, a common challenge with decision trees, and the efficiency in handling complex datasets were pivotal factors in its preference over alternative models. Notably, by encoding attribute values without one-hot encoding, we aimed to enhance model understanding and achieve superior accuracy without overfitting, a strategic choice in optimizing performance.
- **Sub-crop**: In our sub-crop module, we utilize a refined dataset containing information on 22 main crops and their corresponding optimal sub-crops. This dataset which was carefully compiled from an intensive research work contains the data on the environmental factors like nitrogen (N), phosphorus (P), potassium (K), rainfall, soil pH, and humidity.
  To determine the best sub-crop for a given plot of land, we adopt a method inspired by the K-nearest neighbors (KNN) algorithm. We position the sub-crops within a multi-dimensional feature space, with each sub-crop represented as a point. By calculating the distance between the land's environmental features and those of each sub-crop, we identify the sub-crop that is closest to the land in this feature space.
  This method will allow us to guide farmers to make accurate recommendations which take into account their specific land conditions and maximize agricultural productivity. We enable farmers to make informed decisions that boost both productivity and sustainability .
- **Real-time Data-set Retrieval** In the agricultural web application system, the web driver plays a crucial role in automating the retrieval of real-time market data, which is essential for the Market Price Prediction module. Here is how the web driver is implemented and operates within the system:
  Initialization:
  The web driver is initialized using a suitable driver for the browser (e.g., ChromeDriver for Google Chrome, GeckoDriver for Firefox). The driver path and desired capabilities are specified to ensure the web driver can control the browser effectively. Configuration:
  Browser-specific options are configured to optimize the performance of the web scraping tasks. This might include setting the browser to run in headless mode (without a GUI) to conserve resources and disabling image loading to speed up page load times. Navigating to the Target Website:
  The web driver navigates to the target website, such as a government agricultural market database, by sending a request to the website's URL. Session Management:
  Once the target page is loaded, the web driver manages sessions, cookies, and handles any pop-ups or disclaimers that may interfere with data access. Data Extraction:
  The web driver locates the HTML elements that contain the market data using selectors such as IDs, class names, or XPaths. It then extracts the required data, such as crop names, market rates, and dates. Handling Pagination and Navigation:
  If the data spans multiple pages, the web driver navigates through pagination by clicking on the appropriate buttons or links. It waits for new pages to load and continues the data extraction process until all necessary data is collected. Data Storage:
  Extracted data is stored in a structured format, such as CSV or a database, for further processing by the Market Price Prediction module.
- **Market Model**: A Random Forest model proved to be the most accurate with the prediction accuracy of 87% on the carefully crafted dataset incorporating detailed crop market data. This dataset encompasses essential features such as District Name, Market Name, Commodity, Variety, and Grade, with the inclusion of three target variables: the minimum, maximum, and modal prices of crops.
  Our approach entailed the development of a multi-tiered prediction system. In the beginning, we trained three different models. The first model used the independent features to predict the minimum price. After that, we added the predicted minimum price to the feature set and trained the second model to predict the maximum price. In the end, both the minimum and maximum prices were included in the third model to predict the modal price.
  The iterative approach taught us that providing a range (minimum and maximum) for the model to train on significantly improved the prediction accuracy. As a result, we selected this strategy of three-level prediction.
  During inference, user input goes through the first model

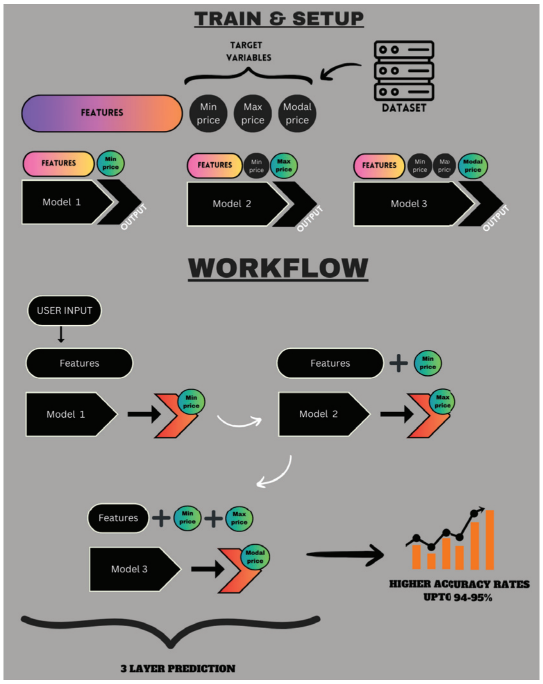

Fig. 2. Market Model

This is the architecture diagram of the workflow of the market model.

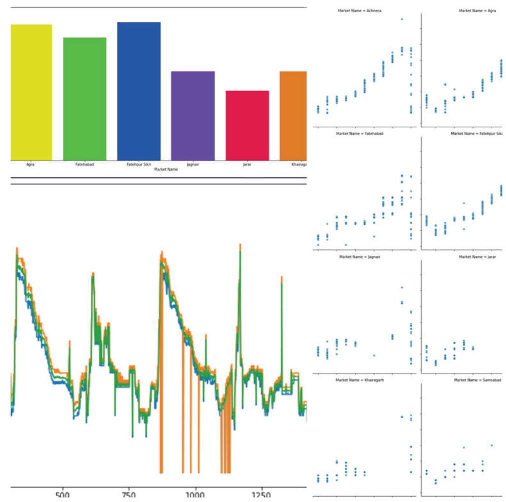

Fig. 3. Market Analysis

The above diagram illustrates the insights derived from the crop data automatically downloaded from the Agmarknet government website.

initially to generate the minimum price prediction. After that, the predicted minimum price is added to the feature set and passed to the second model, which generates the maximum price. At last, combining the highest predicted price with the input data, we deliver this enhanced dataset to the third model which gives the median price prediction.

This advanced methodology continuously provides outstanding results with prediction accuracy of up to 95%.

The steps prior to model training:

1) Data Loading: The real-time dataset downloaded is accessed as a pandas data frame.
2) Data Preprocessing: we make sure we manage missing values, transform date columns, and extract relevant features reflecting market trends and seasonal variations.
3) Encoding Categorical Variables: To maximize the learning quality of the model we tried to enrich the features by one hot encoding suitable features keeping ion mind we don't lean towards over fitting.

E. Model Deployment

1) Main Crop Model: Seamlessly incorporates the trained main crop prediction model into the overall crop prediction system. Ensures compatibility with soil data, weather data allowing for combined predictions to determine main crop suitability.

2) Sub-Crop Model: Effectively integrates the trained sub-crop recommendation model into the broader crop prediction framework and the result to market data to predict the market price of these crops.

3) Weather Model: Seamlessly deploys the trained weather information model within the crop prediction system. Ensures smooth compatibility with soil data for comprehensive predictions of crop suitability based on weather conditions.

4) Market Model: Integrates the trained market price prediction model seamlessly into the crop prediction system. Facilitates compatibility with various parameters, offering insights into predicted crop prices based on market and district.

F. Monitoring And Maintenance

Saving trained models to ensure future scalability and quick deployment in real- world scenarios. Considering to incorporate documentation and error-handling mechanisms for long-term maintainability. It includes a dependency file that lets users to download all the dependency files to run the web application in their system. Regularly update the weather prediction model to incorporate new data and enhance forecasting accuracy. Monitoring model performance over time to identify potential degradation and initiate updates as needed. Maintains thorough documentation on data sources, preprocessing steps, and model configurations for future reference.

G. Results

1) Main Crop Prediction: Our crop prediction model, developed with Kaggle's dataset and leveraging Random Forest, excels in comparison to alternative algorithms. Through a comprehensive evaluation against SVM , KNN, Decision Tree, and Naive Bayes, Random Forest demonstrates an exceptional accuracy of 99.5%. This solidifies its standing as the preferred choice for capturing complex relationships and providing precise main crop predictions, thereby proving invaluable for informed agricultural decision-making.

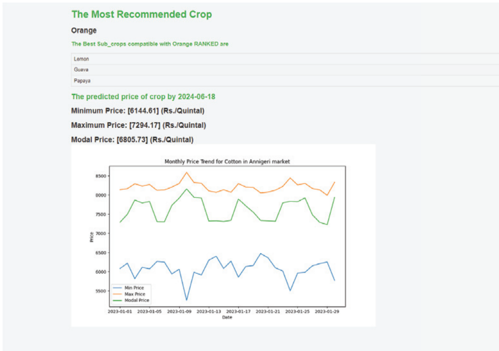

Fig. 4. Final Results

The final page of our web application which summarizes all the conclusive data.

2) Weather API: In this script, we retrieve weather data for a specified location (latitude, longitude) through the Visual Crossing Web Services API, covering the previous 30 days. We accumulate temperature, humidity, and rainfall values from the CSV response. If data is available, we compute the average temperature, humidity, and monthly rainfall, rounding to two decimal places, and present the results as a dictionary.

3) Market Model: We initially trained a Random Forest model using all input attributes to predict three output attributes—minimum, maximum, and modal prices of the crop as according to our research random forest gave the best result when it came to our dataset. This comprehensive model achieved an accuracy of 88%. Here we tried to introduce an intermediate step. We individually predict the max price using a separate Random Forest model and add it as an input feature along with the other attributes for predicting the modal price. This process leads to an accuracy improvement to 90%. Upon careful consideration, we recognized that individually forecasting one output attribute and incorporating it into an- other model as an additional input feature, alongside the existing attributes, could enhance accuracy. This strategic approach, leveraging a highly correlated feature, prompted us to implement this method, including the max price prediction after the min. The outcome was a substantial accuracy improvement in modal price prediction, reaching an impressive 94%.

H. Website Implementation

Users input their location, and if not provided, the system automatically extracts the current location. Subsequently, the system initiates an API call to retrieve weather information, obtaining essential weather parameters for the crop prediction model. Crop model: User-entered soil parameters are combined with weather data in the Crop Model to predict the main crop. Subsequently, the Sub-crop Model suggests suitable sub-crops based on the main crop prediction. Users also receive market insights as the system loads relevant datasets from government websites, predicting prices for all crops based on the specified date and state.

IV. Conclusion And Future Work

Using the above mentioned models alongside real-time dataset retrieval and precise location information, we've successfully created a centralized platform for farmers.This platform serves as a one-stop destination, empowering farmers to make informed decisions about crop selection well in advance while considering prevailing market conditions. They not only facilitate the cultivation of diverse crops, benefiting the land, but also assist farmers in selecting crops with high demand, translating to increased earnings with reduced risks. For market models, our system has an accuracy of 94%, and for the main

crop model, an accuracy of 99.5%, both achieved with Random Forest as the ML model. In the end, making these predictions better, localizing them, keeping them updated, making them sustainable, and making them easy to use are all crucial. It's all about helping farmers make better decisions and grow crops more successfully.Integrating insights on efficient irrigation methods and fertilizer suggestions further enriches the system, offering comprehensive support to farmers for informed decision-making and successful crop cultivation.

References

[1] Sonal Jain, Dharavath Ramesh. "Machine Learning convergence for weather-based crop selection." IEEE (2020).
[2] Ishita Ghutake, Ritesh Verma, Rohit Chaudhari, and Vidhate Amarsinh. "An intelligent Crop Price Prediction using a suitable Machine Learning Algorithm." ITM Web of Conferences 40, 03040 (2021).
[3] S. Geetha and R. Maniyosai. "An Analysis of Cropping Pattern and Crop Combination Regions of Thiruvarur District (2011-2016), Tamil Nadu, India." DOI: 10.36224/ijes.120404.
[4] Anjana, Aishwarya Kedlaya K, Aysha Sana, B Apoorva Bhat, Sharath Kumar, Nagaraj Bhat. "An efficient algorithm for predicting crop using historical data and pattern matching technique." Global Transitions Proceedings volume 2, issue 2 (2021).
[5] Daneshwari Modi, Ashok V. Sutagundar, Vijayalaxmi Yalavigi, and Anupama Aravatagimath. "Crop recommendation using a machine learning algorithm." (ISCON), Oct 22-23, 2021.
[6] https://agmarknet.gov.in
[7] https://data.gov.in/search?title=market
[8] https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset
[9] https://imdpune.gov.in
[10] icar.org.in
[11] kvk.icar.gov.in