INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD)
© 2026 IJNRD | Volume 11, Issue 5, May 2026 | ISSN: 2456-4184 | IJNRD.ORG

MARKET-DRIVENCROP RECOMMENDATION SYSTEM USING PUBLIC DATASETS

1st Rekha V

Department Of Advanced Computing And Analytics Vels

Institute Of Science, Technology & Advanced Studies

Chennai, Tamil Nadu rekhaviswa2004@gmail.Com

2nd Dr.R.Balamurugan

Department Of Advanced Computing And Analytics Vels

Institute Of Science, Technology & Advanced Studies

Chennai, Tamil Nadu bmurugan.scs@vistas.ac.in

Abstract- Crop selection is critical to determining the profitability of a given farm. Many farmers use either traditional knowledge or general advice when selecting their crops, which do not guarantee they will receive the highest profits. This project proposes a machine learning algorithm to sort and analyze the profits from several different crops by applying several types of information (soil content and nutrient levels, climatic conditions, and market conditions) to create a Random Forest model to predict the profitability of a set of crops.

The proposed system utilizes the output from its model to provide suggestions as to which crop will yield the highest profit. The analysis performed on the data indicates that by combining data from agriculture with data from the market, agricultural producers can enhance their decision-making capabilities.

This system allows producers to decrease their exposure to risk, prevent/protect against potential loss, and increase their income through their crop selection process. It is a well- supported and practical solution to modern-day problems in agriculture.

Keywords—*text to handwriting, handwriting generation, LSTM, TensorFlow, Flask, image synthesis, deep learning,*

I. INTRODUCTION

Agriculture is essential for providing food to humans and supporting economies in various regions of the world, especially India. A significant hurdle that farmers encounter is making crop selections at the appropriate time. Many variables play into how a farmer chooses the crops to grow; they include the state of the soil, weather, water supply, and need of the marketplace.

Farmers often base their crop selections off prior experiences and/or well-established farm routines. Regrettably, this approach does not consider how much time has passed since the most recent update in the farming climate has occurred nor how much has changed since the previous day in the marketplaces. These factors lead to a disconnect between the amount of crops produced and the profit derived from those crops. Technology has exploded over the past 2-3 decades, and machine learning has opened new opportunities for analysing large datasets and discovering patterns that are not easily found through a manual process. With machine learning's ability to access historic data for farming and the marketplace, it can be possible to estimate through predicted models that specify which crops will be most productive under the current agricultural settings.

In this proposal, a machine learning based system will be developed for farmers through the use of historical data to recommend crop selections based on several points of interest to the farmer. Ultimately, this machine learning application will help farmers determine which crops to grow, resulting in an increase in quantity and value of crops to the farmer.

II. LITERATURE SURVEY

There have been many research efforts that have looked into ways of using machine learning, natural language processing, and data mining techniques to identify and mitigate fraudulent activities. One of the first methods for detecting fraudulent online content was through the use of text classification algorithms. Researchers have applied different machine learning approaches such as Naive Bayes, Support Vector Machines (SVMs), and Logistic Regression in order to uncover trends within the text. The algorithms assess various features within the job descriptions or job advertisements such as keywords, sentence structure, or context. The results of studies conducted have shown that machine learning classification models can successfully identify fraudulent patterns from large datasets, thus making them suitable for detection of fraudulent activity. However, while machine learning models are generally accurate in their predictions, many traditional models in Artificial Intelligence operate in a manner known as "black boxes" (where the model does not provide any insight into how the model arrived at a prediction). Because of this, the users of these systems may have difficulty trusting the decisions made by the system. To address this limitation, researchers have developed Explainable Artificial Intelligence (XAI) as a means to provide information on model predictions.

III. PROPOSED METHODOLOGY

A Public Dataset-Based Market-Driven Crop Recommendation System

This research will provide an innovative and data-driven system to assist farmers in determining which crops are most profitable based on market conditions and environmental factors. Traditional systems of selecting crops primarily rely on a combination of farmer's experience, seasonal and/or general agricultural practices. Such methods of selecting crops do not take into account the dynamic nature of variables like market price fluctuations, environmental changes, and variability in soil fertility that can result in farmers experiencing reduced yields and/or significant financial losses.

To overcome these challenges, a system is proposed that combines publicly available agricultural datasets with market price data; soil nutrient levels; climate data (temperature, humidity, rainfall); and cost of production data. The proposed system will use machine learning techniques to analyze historical trends in the above datasets and use that analysis to predict expected profits for different crop types under specific environmental conditions.

This research will specifically use the Random Forest Regression model to provide a quantified estimate of crop profitability by using multiple predictor variables, such as nitrogen, phosphorus, and potassium fertility levels; soil pH; soil moisture; temperature; rainfall; humidity; market demand; and the cost of producing a given crop. The Random Forest Regression model will be trained and validated using preprocessed data to guarantee maximum confidence in the accuracy and reliability of any predicted results.

Users will input soil/real-time weather condition data into the proposed system to predict profitability for each crop in the database. The proposed system will rank crops by predicted profitability and provide recommendations to plant (3 crops) so that farmers may make economically sound decisions based on their location and minimize risk associated with uncertainty in agriculture markets.

Since the system uses free public datasets, it will be scalable/cost-effective to implement and will enable farmers to use the system without having to install expensive sensors in their fields or homes.

The proposed work will help to support smart farming practices, improved farmer income, and sustainable agriculture through the integration of agricultural knowledge and technology/data science.

IV VARIOUS MODULES

Module 1: Collecting and Preparing Data

This module helps to collect and prepare the required data (both agricultural and market) needed to develop the crop recommendation model. Public datasets are sourced from government agriculture websites, weather databases and other open data sources. The datasets consist of a variety of attributes including soil nutrient levels (Nitrogen, Phosphorous, & Potassium), soil pH, moisture level, temperature, rainfall, humidity, crop yield, production cost, and market pricing trends.

Data cleaning is performed after data collection to enhance its quality and usability. Cleaning involves (but is not limited to): fixing any missing values, removing duplicate records, fixing any inconsistencies in the data (e.g., incorrect spelling), converting categorical variables (e.g., crop type, soil type, etc.) to numeric values using label encoding. Once all data has been cleaned, it will be in an appropriate structure and format for use in machine learning.

Module 2 : Developing a Machine Learning Model

The goal of this module is to create a predictive model that can estimate the expected profit for crops grown under various environmental and market conditions. The data that has been cleaned will be split into two separate data sets: one for training purposes, and one for testing purposes so that we can assess how well our model performs. We will use a Random Forest Regression algorithm because it is able to process complicated and many-to-many relationships between the input variables and produce highly accurate predictions. We will utilize many different aspects of the data to train our model, including, but not limited to: several soil nutrients, the climate in which each crop is grown, crop yield, market price, and production cost. This additional information will be useful in determining the price of each crop. The Random Forest algorithm will develop a pattern based on the data provided from previous years through the development of numerous different decision trees in order to provide a greater reliability of prediction.

Module 3: crop recommendations. A user provides input parameters, such as soil characteristics, weather conditions, expected yield, and market information. The trained model will predict expected profit based on these inputs for each crop in the dataset.

Crops will be ranked from highest to lowest expected profit based on their predicted expected profits.

V. ARCHITECTURE DIAGRAM

The diagram below shows how the different parts of the system connect to each other.

Figure 1: Crop Recommendation System

VI VARIOUS PHASES AND METHODOLOGIES

Market-Driven Crop Assortment System's construction was done in a logical and concise way to guarantee correct outcomes, resource use, and application of the system in practice. The overall project was divided into different phases (steps) of focus for each of the major functions performed.

1. Problem Identification

The initial/main focus of the first phase was to gather information about the problems of agriculture and farmers; their crops grow according to traditional practices based on Crops agribusiness (does not include comparisons of crop-growing practices is based on what would be categorically defined this as create lower profit margins), therefore providing the solution to the preceding problem by

recommending a process for determining which kind of crops will produce the highest level of economic returns

2. Data Collection

In the second/final phase for data collection, all data that could be used for determinations made for farmer had been collected. The agricultural data used for the database includes

- Possible SOIL pH level, nitrogen, phosphate, potassium,
- Moisture (water) in their soil,
- Weather such as Temperature, Rainfall, Humidity,
- Market Information such as the price of those crops produced by farmers, the demand for that space, and so on.

The above-mentioned agricultural data serves as the primary base/point of assist to the recommendations.

3. Pre-Processing Data

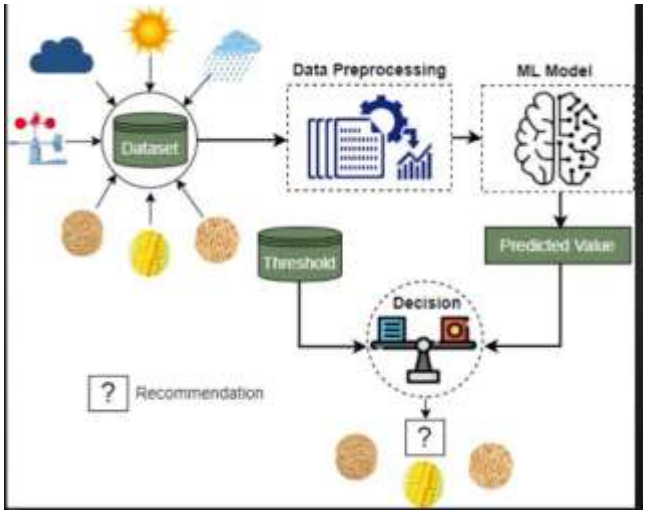
Raw data that is collected from many sources are, too frequently, Missing Data and Missing Consistency, so processing includes data-cleaning measures as follows:

- To handle Missing Data
- To remove duplicate records
- To change from Categorical-to-Numerical Data,
- To normalize the data to better performable action.

This phase assures that the data collection is finalized and prepared for the Data Analysis Step have had passed.

4. Selecting and Developing Features

Crop and market profit variables were identified as the most important variables. Additional features were developed to help improve the accuracy of predictions. This allows the model to focus using



only the most important inputs.

5. Building Prediction Models

Utilising machine learning algorithms to create prediction models, different methods such as regression and ensembles (i.e. random forests) were used to determine patterns and predict crops from input conditions.

6. Training and Testing Prediction Models

In order to test the model's ability to generalise to new data sets, the data was divided into training and testing sets:

- Training Data: Used to train the model
- Testing Data: Used to test the accuracy of the model

7. Assessing Prediction Model Performance

To help determine which of the trained models had the highest accuracy, several metric assessments were used to evaluate accuracy:

- Accuracy
- Mean Squared Error (MSE)
- R-squared Score

8. Implementing the Prediction Model

The prediction model has been developed and implemented within a user-friendly system where a user enters various parameters; their crop plant recommendations will be generated by the prediction model.

9. Result analysis

An analysis of the system output was performed in order to guarantee that the suggestions provided by the system are feasible and of value. The system correctly predicted the maximum crop yields and profits for farmers.10.

10. Deployment and Future Enhancement Applying the System And Future Improvements The last phase is to prepare the system for the real world.

- Connecting to real-time weather data through the use of APIs
- Developing a mobile application for the system

- Adding additional regional database sets to the system

VII. PSEUDOCODE AND IMPLEMENTATION

A. Pseudocode

```
# MODEL
model = RandomForestRegressor(n_estimators=100)

# TRAIN MODEL
model.fit(X_train, y_train)

# EVALUATION
y_pred = model.predict(X_test)

print ("R2 Score:", r2_score (y_test, y_pred))
print ("MAE:", mean_absolute_error(y_test, y_pred))

# RECOMMENDATION
for crop in crop_list:
profit =
model.predict(input_data_with_cr
op) store profit

best_crop =

crop with max

profit # SAVE

MODEL
joblib.dump(model,"model.pkl")
```

B. Implementation Notes

A system is being developed that will recommend crops based on both market and environmental data. The development of the system is done in Python, and will use many machine learning and data analysis libraries. The development process will be completed in a step-by-step manner to ensure that the system provides good crops to recommend.

1. Programming language

- Python is used as the primary programming language for the system because it is relatively easy to learn and use. It has been around for a while and is used by many people in the field of data science and machine learning.

2. Libraries/Frameworks used

- Pandas will be the library that is used to load, clean, and manage the dataset.
- NumPy will be used to perform numerical mathematics and manipulate arrays.
- Scikit-learn will be used to create and train the machine learning model.
- Matplotlib will be used to create graphs and visualizations to make it easier to understand the analysis of the data.
- Joblib will be used to save the trained machine learning model when it has been trained, and then loaded again to be reused without the need to retrain it.

3. Implementation Steps

Step 1, Loading the Data Set
The data set is loaded from an Excel document in the Pandas data library. It has a lot of useful information about soil nutrient levels, the weather, and prices in the market.
Step 2, Data Preprocessing
The data set must also go through some kind of data cleansing process by identifying missing values, duplicates, and inconsistencies for better results from the modelling process. Step 3, Feature Engineering
The final features used to train the model include, for example, soil types, average temperature, average amount of rainfall, and prices in the market. Categorical variables such as crop type and soil type are then encoded as numbers to be usable as features.
Step 4, Model Training
We will use the Random Forest algorithm in our model and separate our data set into training and test data sets to allow the model to learn to properly predict.
Step 5, Model Evaluation
Finally, we will evaluate our model using various performance metrics such as R^2 Score and Mean Absolute Error (MAE) to determine how accurate our model's predictions were.
Step 6, Recommendation
This trained model will predict the expected profit for various crops; therefore, the recommendation system will recommend to users what crop they should grow for the highest return.
Step 7, Data Visualization
The use of Matplotlib to generate graphs or charts from the data will allow users to visualize trends (e.g., increases in crop prices or crop yields) and thus help them make more informed decisions about what crops to grow.
Step 8, Saving the Model
Finally, we will save the model we've trained using Joblib, allowing for later predictions through this model without requiring retraining.

X. SCREENSHOTS, CHARTS, AND GRAPHS

The figures below show the web interface and sample outputs generated during testing.

1. Future Importance

INPUT INPUT & OUTPUT

- Soil: Loamy
- Rainfall: 850 mm
- Temperature: 28°C
- Market Price: 2000

OUTPUT

Output:

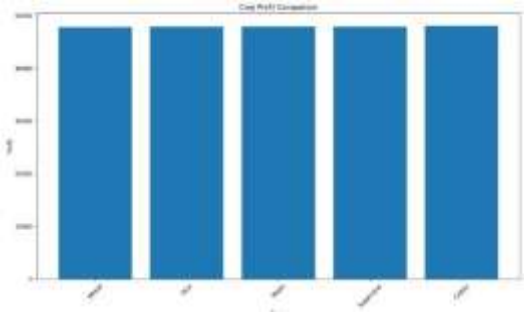
Best Crop:
Rice Expected Profit: 82000

VIII. RESULT AND DISCUSSION

The system successfully predicts the most profitable crop based on input conditions such as soil type, rainfall, temperature, and market price. For example, when the input values are loamy soil, rainfall of 850 mm, temperature of 28°C, and market price of 2000, the system recommends Rice with a high expected profit.

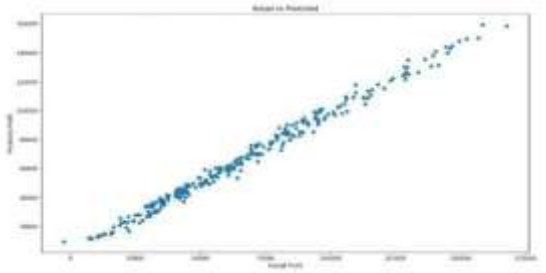
The model shows high accuracy with an R^2 value close to 0.98, which indicates strong performance. It is observed that factors like rainfall and market price play a major role in determining profit. The system also works well for different districts and seasons.

These results prove that combining agricultural and market data can significantly improve crop selection and decision- making.

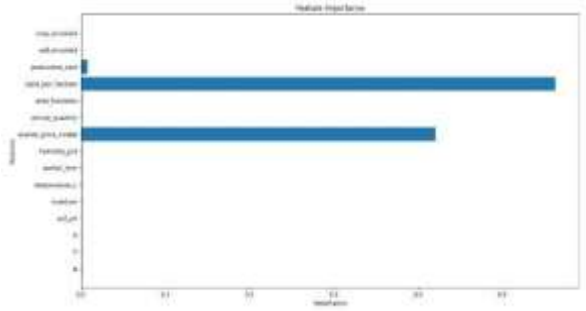


3. Crop Profit Comparision

2.



4. Output



XI. CONCLUSION

This project successfully demonstrates how machine learning can be applied in agriculture to improve crop selection. By considering both environmental and market factors, the system provides accurate and practical recommendations.

The use of the Random Forest model ensures reliable predictions, and the inclusion of market data makes the system more realistic and useful. It helps farmers choose crops that can give better profit, reducing financial risk and improving productivity.

Although the system performs well, it can be further improved by integrating real-time data such as live weather updates and market prices. Future development can also include mobile or web applications for easier access.

Overall, this project highlights the importance of data- driven approaches in agriculture and shows how technology can support farmers in making better decisions.

REFERENCES

[1] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[2] R. Sujatha and P. Isakki, "A Study on Crop Prediction using Machine Learning," *International Journal of Engineering Research & Technology (IJERT)*, vol. 9, no. 5, pp. 123–126, 2020.

[3] K. Ramesh and D. V. Vishnu, "Crop Recommendation System using Machine Learning," *International Journal of Advanced Research in Computer Science*, vol. 10, no. 2, pp. 45–50, 2019.

[4] Food and Agriculture Organization, "The State of Food and Agriculture," FAO, Rome, Italy, 2020.

[5] Ministry of Agriculture & Farmers Welfare, "Agmarknet: Agricultural Marketing Information System," Government of India. [Online]. Available: https://agmarknet.gov.in

[6] India Meteorological Department, "Weather Data Services," Government of India. [Online]. Available: https://mausam.imd.gov.in

[7] Scikit-learn, "Random Forest Regressor Documentation." [Online]. Available: https://scikit-learn.org

[8] Pandas, "Python Data Analysis Library." [Online]. Available: https://pandas.pydata.org

[9] NumPy, "Numerical Computing in Python." [Online].
Available: https://numpy.org

[10] Matplotlib, "Visualization with Python." [Online]. Available: https://matplotlib.org