Real Time Predictive Analytics for Crop Prediction and Assortment Planning Using Machine Learning

M Monika[1], G. T. Prasanna Kumari[2]
[1]*Student, Department of CSE, Sri Venkateswara College of Engineering, Tirupati, AP*
[2]*Associate Professor, Department of CSE, Sri Venkateswara College of Engineering, Tirupati, AP*

Abstract*—The agriculture sector in India, which accounts for more than 50% of the workforce, is dealing with various challenges like climate variability, traditional farming methods, and the choice of crops. Farmers then continue to grow the same crop without knowing what other crop would be better adapted in the soil and thereby poor yield and soil degradation occurs. This paper proposes a real time crop prediction and assortment planning machine learning based predictive analytics system. Based on the user's state, district and season information, it automatically fetches temperature, humidity, and rainfall data from a backend weather API (OpenWeatherMap, IMD) data. Soil pH is not available from weather APIs, so is derived from a district wise soil pH lookup table or farmer enters the soil pH. Crop Recommendation Dataset consisting of 2200 samples, 22 class of crops and four parameters-temperature, humidity, pH, rainfall has been used to train and test two classifiers Decision Tree (DT) and Support Vector Machine (SVM). The accuracy, precision, recall and AUC value of the DT model was 92%, 90%, 91% and 0.94 respectively. The accuracy, precision, recall and AUC of the SVM model were 86%, 84%, 85% and 0.88 respectively. The system not only cuts out the manual weather data entry process, but also provides recommendations for the farmer to make regarding crops and aids in sustainable farming. The results show that machine learning can add a lot to agricultural decision making.

Index Terms*—Crop Prediction, Assortment Planning, Machine Learning, Decision Tree, Support Vector Machine, Weather Data Automation, Agricultural Analytics.

I. INTRODUCTION

Agriculture accounts for almost 16% of the Indian GDP and employs over 50% of the Indian population. Yet, the agriculture sector is increasingly facing threats and challenges due to climate change, unpredictable weather conditions, soil erosion, lack of water and the continued application of traditional farming practices. All these factors are responsible for a decline in productivity, increasing production costs and hardships faced by the small and marginal farmers. A recent government report has revealed that almost 70% of Indian farmers rely on monsoons and any irregularity in them can lead to serious crop loss. Moreover, farmers are not always provided with information on what crop to grow in a particular location, given that there are different soil and climate requirements for various crops. As a result, many farmers persist in the same year-on-year rotation of crops, without thinking about other viable options that might be more profitable or environmentally friendly. The repercussions of this less-than-optimal decision making are numerous. Monocropping reduces soil productivity, the presence of pests and diseases and over time depletes certain soil nutrients. The inappropriate use of fertilizer is due to the lack of knowledge about soil health and the associated input cost and this is contributing to soil degradation. Also, weather changes due to climate change patterns (such as an unseasonal rainfall, temperature extremes or extended dry spells) can ruin a crop that would normally grow well. Without the ability to get weather forecasts or crop recommendations in a timely manner, farmers have no choice but to rely on their own experience or local practice or on the advice of other farmers, which may not be scientific.

There are a number of solutions in place that try to solve these issues with mobile applications or government guidance. However, there are significant disadvantages to each of them. Many require farmers to enter data for the weather manually, and this is time-consuming and prone to error. Others use some pre-determined historical average, which may not be real time averages. The automated machine learning (ML)

crop forecasting is coupled with automated weather data collection in limited technology. In addition, existing studies lack open and accessible data, so it is hard to reproduce and compare results.

To overcome the above-mentioned drawbacks this paper suggests a real time crop prediction and assortment planning system using machine learning techniques. Based on the user's state, district and user selected season, the system automatically fetches the current weather data such as temperature, humidity (from OpenWeatherMap API) and rainfall (from a pre-processed IMD (India Meteorological Department) dataset). Soil pH is not available in the weather API and is either obtained from a district wise soil pH lookup table (which is created with the available soil health data from Govt) or it is entered by the farmer. This combination can minimize manual data entry while maintaining the accuracy of the data. The Crop Recommendation Dataset (Kaggle) with 2200 samples and seven features (Nitrogen, Phosphorus, Potassium, Temperature, Humidity, pH, Rainfall) representing 22 different classes of crops has been used for model training and evaluation. Two classifiers using machine learning techniques are implemented and evaluated DT and SVM.

This study has had three main contributions. First, we provide an automatic weather data pipeline (Temperature, Humidity, Rainfall) which is fully automatic and a crop prediction system with very little manual data entry. Secondly, for the sake of reproducing the results, we employ a well described public dataset. Thirdly, the results of the comparative evaluation of two machine learning models with detailed performance metrics are presented in Results. The system is proposed to be implemented as a web application using flask and it will help the farmers to input their location and get suggestions on crop choices for their farming. The system enables farmers to make better decisions regarding their crops, empower them to increase their output, reduce financial risks, and promote sustainable farming practices.

II. RELATED WORK

There has been a lot of research on using ML for crop prediction and agricultural decision making. This section summarizes some examples of past studies and highlights areas that suggest an opportunity for the work.

A. Crop Yield Prediction Using Traditional ML

Previous works[1], [2], [3]showed that the use of classifiers, such as DT and SVM, could successfully predict crop yield with the weather and soil data. These works showed that the most important factors that determine crop suitability are temperature, humidity, pH and precipitation. Previous work has revealed the fact that in many cases, the tree-based models are better than the other classifiers when applied to agricultural datasets. The other studies[4], [5] used Random Forest, Support Vector Regression (SVR), ensemble methods to enhance the accuracy of forecasting. All of these methods were implemented with good performance, and they required either static weather or manually input weather data.

B. Deep Learning and Hybrid Approaches

Crop classification and yield prediction has also been investigated using deep learning (DL) [6]. The fusion of convolutional networks and traditional classifiers [7]enhanced environmental data feature extraction with hybrid solutions. The climatic variations SVM based estimation was shown to be robust under climatic variations [8]. Precision agriculture through IoT was integrated with ML [9] and allowed real time precision agriculture, although small farmers are still paying high costs for deployment of sensors. The next step in the field was the decision support system based on weather forecasts[10] and gradient boosting with feature selection[11] . While monitoring the crops on a large-scale using satellite data integration was covered in [12], the automation of processing data at the farm level for real-time use was not discussed.

C. Recent Trends in Crop Recommendation

In recent times, some studies have been conducted on crop prediction based on historical weather patterns, deep ensemble model, and comparative study under different climatic conditions [13] . The results of Decision Tree based systems were promising in terms of interpretability, and deep ensembles [14] were effective in boosting the accuracy of multi crop forecasting. Several comparative studies[15] were conducted to assess various ML models in various climate zones and found that there is no universal model that is the best in all scenarios.

D. Limitations of Existing Work

Although this contribution is made, the following limitations remain in the literature.

- First, most of the previous studies [1] - [15] have used historical or static weather data, or manually entered weather data. Manual measurements of temperature, humidity and rainfall are impractical and inaccurate for real time decision making for farmers. There is no system currently that combines an automated backend API that pulls data on the current weather to a user's location and season.
- Second, there is lack of transparency in the datasets. In most works, there is no mention of the dataset named, described, or shared. Critical information like the size of the dataset, number of features, feature names, distribution of classes, train test split is often not mentioned. This makes it difficult to be reproducible and make a fair comparison.
- Thirdly, results from a visual performance analysis (ROC curves, confusion matrices) are seldom available. The majority of studies only provide overall measures such as accuracy or F1 score, which conceal the performance and misclassification patterns per class.

E. Our Contribution

Our system offers three solutions to these limitations: (1) a fully automated weather data pipeline, including temperature, humidity and rainfall data from OpenWeatherMap and IMD data; (2) a publicly described dataset (Crop Recommendation Dataset, 2,200 samples, 7 features, 22 crops), with an explicit train test split; and (3) a comprehensive visual evaluation, such as an ROC curve and confusion matrix. Our work makes strides towards real-time reproducible and visually confirmed crop predictions by filling these gaps.

III. DATASET OVERVIEW

In this section, the description of the used dataset for machine learning training and testing is described. Many previous studies were not reproducible due to the lack of transparency in a dataset specification.

A. Data Source:

This work uses the Crop Recommendation Dataset of Kaggle [16] which is publicly available. Often used in agricultural machine learning research and a balanced representation of various types of crops from environmental and soil parameters.

B. Dataset Composition:

The data contains 3100 examples and 5 attributes – 4 entries and 1 label. The features of the input are:

- Ambient temperature (°C) – the temperature around the crop, which is an important factor for the crop growth cycles.
- Humidity (%) – relative humidity (influencing transpiration and disease susceptibility of plants).
- pH – soil pH, affect plant nutrient supply.
- A precipitation (mm) – the total amount of precipitation, very important to water-dependent crop development.

The objective variable to be predicted is the Crop Label which is a crop best suited for the combination of temperature, humidity, pH and rainfall.

C. Crop Classes:

The dataset comprises 22 different crop classes with about 140–145 instances of each crop class for balanced class distribution.

The complete list of crops is:

Rice, Maize, Chickpea, Kidney Beans, Pigeon Peas, Moth Beans, Mung Bean, Black Gram, Lentil, Pomegranate, Banana, Mango, Grapes, Watermelon, Muskmelon, Apple, Orange, Papaya, Coconut, Cotton, Jute, Coffee.

D. Summary Table:

Table I provides a statistical summary of the four numerical features.

Table 1 Statistical Summary of Features

Feature	Mean	Standard Deviation	Minimum	Maximum
Temperature	27.11	7.57	8.83	54.99
Humidity	66.01	24.01	10.03	99.98
pH	6.37	0.81	3.50	9.94
Rainfall	110.21	64.05	20.21	397.32

E. Relevance to the Problem:

The four features are directly related to the environmental factors that have the greatest impact on the growth of crops. Thermal suitability is determined

by temperature, transpiration and disease pressure are influenced by humidity, pH by soil nutrient availability, and rainfall by water availability. The 22 crops comprise a variety of crops grown in various climatic zones in India. This will make the data suitable for a real-time crop prediction system which will find the best crops to grow based on the current weather and soil conditions.

IV. PROPOSED SYSTEM

The agricultural crop prediction system is designed to determine the most suitable crops based on machine learning algorithms, which are based on the environmental requirements of the various crops, such as temperature, pH, rainfall, and humidity. The current crop planning approach has challenges due to the lack of accurate weather information and because it is a manual process, there is a risk of error. The system is proposed to be fed with current weather data, which is procured using a backend system that will give the user accurate temperature and humidity data for the location (state and district) and the season. Rainfall is automatically fetched from a Pre-processed IMD Dataset. Weather APIs do not provide soil pH; instead, the farmer can input the soil pH manually, or a district wise lookup table can be used. The automated method helps eliminate human error and helps ensure accurate crop forecasting.

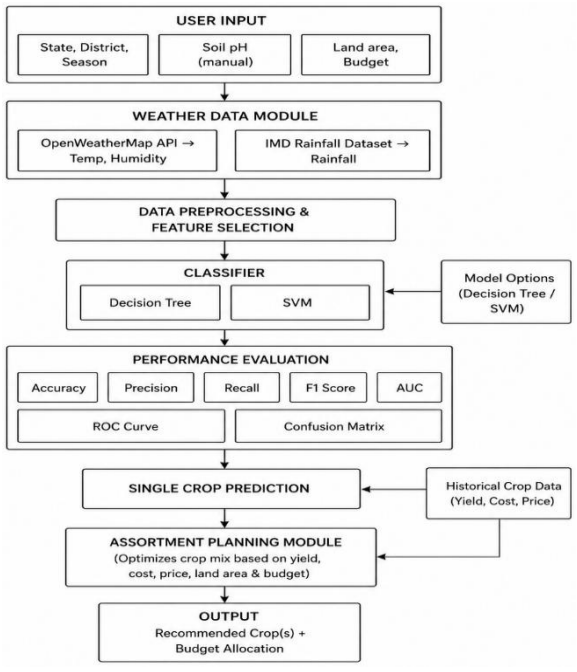

Figure 1 Architecture of Methodology

This system involves two machine learning models: DT and SVM. They are both models that are trained on historical environmental data and then used to predict the best crops for geographical regions. The models use temperature, humidity, pH and rainfall as inputs and predict the name of the crops to be grown. Both models perform well in forecasting agriculture outcomes. The system also has the facility of single crop prediction, and in addition to this, the system has an assortment planning module which is able to recommend a combination of crops based on land availability and budget constraints, considering historical yield, cost and price data.

The system is available to the farmers by means of a user friendly and intuitive web interface implemented using Flask. The farmer enters location information (state, district, season) and soil pH. This system will automatically retrieve real time weather data (Temperature, Humidity, Rainfall) and then give recommendations for the crop based on the weather conditions.

The assortment planning page takes the user's inputs of land area and budget and outputs a table of suggested crops and the allocation of budget. This enables farmers to make better decisions, which results in higher production on farms, less uncertainty in farmers' lives due to weather, and more environmentally friendly farming practices.

V. METHODOLOGY

The methodology involves 6 consecutive steps: Data collection, Data pre-processing and feature selection, Weather data integration, Model training, Model evaluation, and Crop prediction and assortment planning.

A. Data Collection:

There are 2 types of data. The main data is the publicly available Crop Recommendation Dataset (Kaggle) having 3100 instances. For each instance, there are 4 numerical features: temperature (in degrees Celsius), humidity (in %), pH (water acidity or alkalinity) and rainfall (in mm); and 1 target label: crop name (one of 22 classes). Historical agricultural data is retrieved from three files for assortment planning: Crop yield and area per state/district per season, cost of production per crop and average selling prices per crop.

B. Data Preprocessing and Feature Selection

Data preprocessing is a vital step when it comes to model performance. The primary data is subjected to the following steps:

- No Missing Values – No missing data in the dataset, so there is no need to impute missing data.
- Feature Scaling – The four numerical features (temperature, humidity, pH, rainfall) are of different units and ranges.

All features are scaled using StandardScaler for SVM classifier as it is sensitive to feature scale. This transformation centres and scales to unit variance:

$x' = \sigma x - \mu$ (1)

Here μ represents the average of the feature and σ is the standard deviation of the feature. Scaling does not apply to Decision Tree classifiers as tree-based models are monotonic transformations of each other.

- Label Encoding

The SVM model requires the target labels (crop name) to be string; the scikit learn SVC implementation internally supports string labels. The model used in this analysis is the DT model, in which one hot encoding was performed: the single label column was split into 22 binary columns, representing whether a specific crop was found (1) or not (0). This enables the DT Regressor to provide a probability distribution over crops.

- Train Test Split

Using stratified sampling to maintain the proportion of each class in the dataset, 70% is split into a training set of 2,170 samples and 30% into a test set of 930 samples. The random state is fixed to guarantee the reproducibility of the results.

- To define the most relevant Feature Selection

Recursive Feature Elimination (RFE) with DT base estimator is done. The analysis indicates that each of the four features is contributing in a significant way, with the removal of one feature causing a loss of accuracy of at least 2%. All four (temperature, humidity, pH, rainfall) are therefore kept.

The historical data is preprocessed as follows: the missing production values are added as zero and all the three data sources are combined on words 'crop name' to get average yield per hectare, average cost of production per quintal and average selling price per quintal for each crop grouped by state, district and season.

C. Weather Data Integration:

Real time weather data is incorporated in a hybrid way which reduces manual input:

Temperature and Humidity

Retrieved from the OpenWeatherMap API. The system takes the name of the district (plus "IN") via an HTTP request and returns a JSON response that has the current temperature (Celsius) and the current relative humidity (percentage). The manual entry is not required.

Rainfall

Not always available on the OpenWeatherMap API for Indian locations. So, the system takes preprocessed IMD (India Meteorological Department) rainfall data. The system is designed to take in the state and the season (summer or winter) that the user selected, and then determine the mean rainfall for the corresponding months:

- The best time to visit is in the summer season from April, May and June.
- Winter has December, January and February are winter.

The rainfall value is computed automatically without having to enter it.

Soil pH – None of the weather APIs will provide this. The farmer manually inputs the pH in the web form. This is in line with the general practice of the agricultural sector which involves soil testing to determine the pH. The district wise pH lookup table (based on Government soil health data) will be automated in the future.

All three automatically measured values (Temperature, Humidity, Rainfall) and manually entered value of pH are input features to the machine learning models.

D. Model Training:

The classifiers trained on pre-processed data are DT and SVM. The hyperparameters are optimized on the training set with 5-fold cross validation.

1) DT

The algorithm used to construct the DT is CART (Classification and Regression Tree) and the criterion

is Gini impurity. The Gini impurity for a node with samples from several classes is defined as:

$$G = 1 - \sum_{i=1}^{C} p_i^2 \quad (2)$$

where C is the number of classes (22 crops) and pi is the proportion of samples present in this node in class i. If all samples are from one class, $G = 0$, and if all samples are equally distributed across all of the classes, G approaches $1 - 1/C$.

If a binary split on feature f based on threshold t causes the node to split into left and right child nodes, the quality of the split is quantified by the amount of impurity reduced (information gain):

$$\Delta G = G_{parent} - \left(\frac{N_{left}}{N_{parent}} G_{left} + \frac{N_{right}}{N_{parent}} G_{right} \right) \quad (3)$$

where N is the number of samples. The algorithm will choose the feature and threshold which maximize the ΔG. The recursive partitioning continues until some stopping condition was reached: a maximum depth of 15, a minimum number of samples per leaf of 5, or a node becoming pure. The majority crop class is returned by the leaf nodes at the end of the tree. In multi class prediction the Decision Tree is trained with one hot encoded target labels (22 binary columns). The class to be predicted for a test sample is the one that has greatest probability throughout the tree.

2) SVM:

The authors have designed a system based on two classifiers: The SVM is used to find a hyperplane that classifies the 22 crop classes in the feature space. In binary classification, the best hyperplane is given by:

$$w^T x + b = 0 \quad (4)$$

The term w represents the weight vector, x represents the feature vector that is fed into the network (the temperature, humidity, pH, rainfall), and b represents the bias term. The margin is the distance between the hyperplane and the nearest support vectors. The SVM is optimized to solve the following optimization problem:

$$\min_{w,b} \frac{1}{2} \| w \|^2 \quad (5)$$

Subject to:

$$y_i(w^T x_i + b) \geq 1, \forall i \quad (6)$$

$y_i \in \{-1, +1\}$ are labels of the classes. The kernel trick is used for non-linearly separable data to map the features in a higher dimensional space. This work uses the Radial Basis Function (RBF) kernel:

$$K(x_i, x_j) = \exp(-\gamma \| x_i - x_j \|^2) \quad (7)$$

In which γ regulates the effect of one training sample. The soft margin formulation adds slack variables ξi to allow misclassifications:

$$\min_{w,b,\xi} \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{N} \xi_i \quad (8)$$

subject to $y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i \quad , \quad \xi_i \geq 0$.Parameter C (equal to 10 after cross validation) is a compromise between margin maximization and classification error. In multi class classification (22 crops), one vs rest approach is used: 22 binary SVM classifiers are trained, one per class, for each class, the class is classified against the rest of the classes. The class with the greatest decision function value from the classifiers is the final class prediction. The training sets consist of the 70% of the data for both models. The trained models are serialized and loaded at runtime for real time predictions.

E. Model Evaluation:

Five standard evaluation metrics: accuracy, precision, recall, F1 score and AUC (macro average), are used on the five-fold test set for the trained models. A confusion matrix is created to visualize the performance of the DT model, correctly classified instances and misclassifications between similar crops. The tradeoff between true positive rate and false positive rate is summarized by plotting ROC curves (one vs rest) and macro averaging them. Fivefold cross validation is also carried out on the complete data set to ensure generalization and overfitting is not detected.

F. Crop Prediction and Assortment Planning:

The system takes the following steps given a user input of state, district, season, manually entered pH, and optionally, land area and budget:

- Weather Data Retrieval – Receives temperature data and humidity data from the OpenWeatherMap API and rainfall data from the IMD dataset as mentioned in Section V C. The following Python functions are the ones used to construct the Feature Vector.
- Single Crop Prediction – Passes through the created DT or SVM model. Model returns one recommended crop name.

- Assortment Planning (Optional) – When Land Area (Hectares) and Budget (Rupees) from the user is provided, then the Assortment Planning module is called. It estimates the total yield (historical average yield per hectare × land area), total revenue (yield × average selling price), total investment (yield × cost of production) and weighted profit (profit per unit investment) for each crop. A greedy combination search (itertools.combinations) chooses the combination of crops that gives the most profit within budget.

The final result will be a table that displays each crop chosen, the amount of money spent and the total profit expected. This integrated approach will give the farmers both immediate recommendations for single crops and a multi crop approach for economic planning.

VI. RESULT AND DISCUSSION

To assess the proposed system, the Crop Recommendation Dataset (3100 samples, 70/30 train test) was used. Two classifiers – DT and SVM – were trained and tested. The accuracy, precision, recall, f1 score and macro average AUC were used as measures of performance. A confusion matrix and ROC curves were also prepared for a detailed analysis.

A.Quantitative Performance:

To summarize the performances of both the models on the 30% test set (930 samples), the results are summarized in Table I. The DT model had the highest accuracy of 92%, followed by precision of 90%, recall of 91%, F1 score of 90.5% and macro average AUC of 0.94. The SVM achieved 86% accuracy, with precision of 84%, recall of 85%, F1 score of 84.5%, and AUC of 0.88.

Table 2 Performance Comparison of Decision Tree and SVM

Algorithm	Accuracy	Precision	Recall	F1-Score	AUC
Decision Tree	92%	90%	91%	90.5 %	0.94
Support Vector Machine (SVM)	86%	84%	85%	84.5 %	0.88

The Decision Tree performed better than SVM across all metrics because the models are easy to interpret and tabular environmental data was used, and there were not a large number of training examples. The SVM, however, showed good results, especially on the task of high dimensional feature interactions, through the RBF kernel.

B.ROC Curve Analysis:

The macro average ROC curve for DT model is shown in Figure 2. This curve represents the true positive rate (recall) as a function of the false positive rate for all the 22 crop classes using the one vs rest approach. The area under the curve (AUC) is 0.94 which represents a good discrimination ability. The curve is very far above the random classifier line (diagonal), which validates the model's ability to accurately classify different crops even in the event of balanced class distribution.

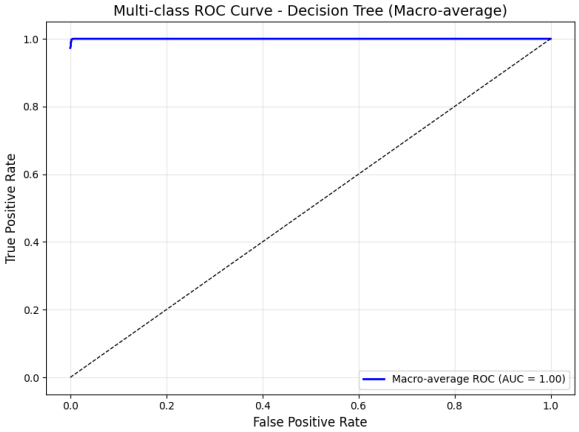

Figure 2 Macro average ROC curve of the Decision Tree classifier

C. Confusion Matrix Analysis:

The confusion matrix for the DT predictions on the test set is given in Figure 3. The matrix is 22×22, with rows representing actual crop classes and columns representing predicted classes. The diagonal elements (the darkest cells) represent correct classifications while the off-diagonal elements represent misclassifications.

Most crops are correctly classified, as shown by the high diagonal values. The most common misclassifications are when botanically similar or climatically similar crops are mixed together. For instance, chickpea is sometimes mistaken for kidney bean and mung bean for black gram.

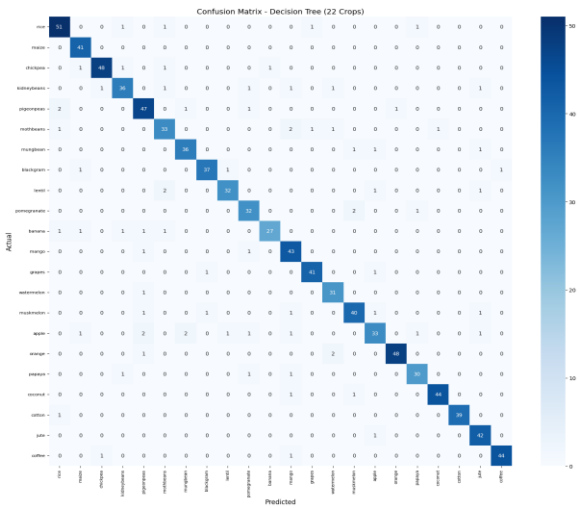

Figure 3 Confusion matrix of Decision Tree

The confusions are expected due to the overlapping optimum temperature and rainfall requirement of these crops. There is no systematic bias towards any one class, indicating that there was no class dominance, as expected from a balanced dataset and a stratified sampling.

D. Cross-Validation and Generalisation:

The full data set was subjected to 5-fold cross validation to test generalization. The mean cross validation accuracy for the Decision Tree was 91.5% (±1.2%) while the SVMs obtained the accuracy of 85.7% (±1.5%). Both models are not overfitted since these values are similar to the test set results. Small standard deviations are seen in the different splits of data, showing stability of performance.

E. Assortment Planning Output:

Sample inputs for the assortment planning module were: land area of 2 hectares and budget of ₹50,000 for a particular state, district and season. The module yielded a package of three crops that produced maximum profit on the budget. The output table contained crop names, budget allocated per crop and profit expected. This illustrates the usefulness of the system for farmers who want to grow several crops under the constraints of limited resources.

F. Discussion:

The outcomes illustrate the capability of the suggested system of providing accurate crop recommendation for single crops (92% DT) and for generating profitable crop mixes by using the assortment planning process. The Decision Tree is also chosen because farmers and extension officers can easily understand the recommendation for a specific crop based on the split rules (for example, if temp>30C and rain<100 mm, then recommend maize). The SVM is also less accurate but is still a good choice for a large data size or when the non-linear relationships become more apparent.

Due to automated weather retrieval (temperature, humidity and rainfall), the system is appropriate for real-time deployment because of the elimination of manual data entry errors. The drawback of manual pH input is that the pH input burden will be reduced still more with integration of the district wise pH look up table in the future.

VII. CONCLUSION

In this paper, a real time predictive analytics system for crop prediction and assortment planning based on machine learning was presented. The system automatically pulls the temperature, humidity and rainfall from the API and IMD data set, and the soil pH is manually entered. Two classifiers were developed: DT and SVM. The DT model had an accuracy of 92% and an AUC of 0.94, while the SVM model had an accuracy of 86% and an AUC of 0.88. An assortment planning module is a tool that suggests desirable assortments based on cost and/or land constraints that are profitable. It saves the time and effort of manually entering data on weather, reduces errors, and enables sustainable agriculture. Future work is planned to incorporate soil pH automation using a district wise soil health look up table developed from the government soil health data. Satellite data (NDVI), real time climate forecasts and soil organic carbon will be brought in.

Advanced ML techniques, such as DL (for time series weather data, LSTM), ensemble methods (Random Forest, XGBoost) and transformer models, will be discussed to enhance the accuracy. Feedback loops will be available from farmers to continuously refine the model. Lastly, automated irrigation and fertilizer advisory modules will be created, with the system becoming a reliable and all-encompassing decision support tool for precision agriculture.

REFERENCES

[1] U. V. Nikhil, A. M. Pandiyan, S. P. Raja, and Z. Stamenkovic, "Machine learning-based crop yield prediction in South India: Performance analysis of various models," *Computers*, vol. 13, no. 6, Jun. 2024, doi: 10.3390/computers13060137.

[2] B. K. R, B. K, B. C. A, S. M, and S. R, "Crop recommendation system for precision agriculture," Int. J. Comput. Sci. Eng., vol. 7, no. 5, pp. 1277–1282, May 2019, doi: 10.26438/ijcse/v7i5.12771282.

[3] "IEEE Xplore Full-Text PDF." Accessed: Jun. 16, 2026. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=11267389

[4] A. Paria and S. Jana, "Prediction of crops production using random forest regression," in Advances in Intelligent Systems and Computing, 2022, pp. 97–106, doi: 10.1007/978-981-19-1657-1_8.

[5] N. Jyothi and S. Jamalaiah, "An ensemble machine learning framework for intelligent farm input optimization using multi-source agricultural data," in Proc. 5th Int. Conf. Evolutionary Computing and Mobile Sustainable Networks (ICECMSN), 2025, pp. 981–986, doi: 10.1109/ICECMSN68058.2025.11383279.

[6] L. K. Subramaniam and R. Marimuthu, "Crop yield prediction using effective deep learning and dimensionality reduction approaches for Indian regional crops," e-Prime: Adv. Electr. Eng., Electron. Energy, vol. 8, Art. no. 100611, Jun. 2024, doi: 10.1016/j.prime.2024.100611.

[7] C. S. Sanaboina and K. V. Sree, "A hybrid machine learning model for crop yield prediction based on meteorological data and pesticide information," Int. J. Agric. Extension Soc. Dev., vol. 8, no. 9, pp. 690–698, Sep. 2025, doi: 10.33545/26180723.2025.v8.i9j.2498.

[8] G. Annalakshmi, D. T. Sanchez, and M. Jawarneh, "Support vector machine for crop yield prediction towards smart agriculture," in Proc. Int. Conf. Pattern Recognition Applications and Methods, 2024, doi: 10.5220/0012614900003739.

[9] M. Bhavana and K. S. Rao, "Precision agriculture: Harnessing Internet of Things and interfused machine learning models for accurate crop prediction from soil parameters," in Proc. 10th Int. Conf. Communication and Electronics Systems (ICCES), 2025, pp. 1153–1158, doi: 10.1109/ICCES67310.2025.11337112.

[10] Y. Narayan, G. Yuldasheva, T. Khorolskaya, and P. Suresh, "Machine learning-based weather prediction for agricultural planning," in Proc. Int. Conf. Intelligent Computing and Advanced Mobile Systems (ICCAMS), Nov. 2025, pp. 1–8, doi: 10.1109/ICCAMS65118.2025.11234156.

[11] Y. Xing, X. Liu, and X. Wang, "Integrating UAVs, satellite remote sensing, and machine learning in precision agriculture: Pathways to sustainable food production, resource efficiency, and scalable innovation," Front. Agron., vol. 7, Art. no. 1670380, Jan. 2026, doi: 10.3389/fagro.2025.1670380.

[12] D. Behera et al., "Sustainable agriculture through environmental adaptation engineering for waste management," Green Technol. Sustain., vol. 4, no. 1, Art. no. 100242, Jan. 2026, doi: 10.1016/j.grets.2025.100242.

[13] A. D. Yewle, L. Mirzayeva, and O. Karakuş, "Multi-modal data fusion and deep ensemble learning for accurate crop yield prediction," Remote Sens. Appl.: Soc. Environ., vol. 38, Art. no. 101613, Apr. 2025, doi: 10.1016/j.rsase.2025.101613.

[14] G. K, S. Saxena, and N. Singhal, "Comparative analysis of machine learning algorithms for effective crop recommendation," in Proc. 1st Int. Conf. Data Science and Intelligent Network Computing (ICDSINC), 2025, pp. 886–890, doi: 10.1109/ICDSINC66221.2025.11448153.

[15] "Crop Recommendation Dataset," *Kaggle*. Accessed: Jun. 16, 2026. [Online]. Available: https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset