Crop Recommendation System using ML

Suniti Ghorai[1], Arohan Gupta[2], Arvind Kumar[3], Piyali Maur*

[1]Suniti Ghorai: Department of Computational Sciences, Brainware University, Kolkata-700125, E-mail: sunitighorai06@gmail.com

[2]Arohan Gupta: Department of Computational Sciences, Brainware University, Kolkata-700125, E-mail: arohangupta56@gmail.com

[3]Arvind Kumar: Department of Computational Sciences, Brainware University, Kolkata-700125, E-mail: arvindkumararyan2001@gmail.com

*Piyali Maur: Corresponding Author, Department of Computational Sciences, Brainware University, Kolkata-700125, E-mail: piyali.maur@gmail.com

ABSTRACT

Agriculture is the world's leading source of food particulars. All the food substances that are essentialize. Agriculture produces vegetables, proteins, and canvases. The carbohydrates give all living beings with energy. growers have great significance in our society. They're the ones who give us food. Presence of Too numerous interceders Middlemen results in the exploitation of both growers and consumers with the mediators offering lower prices to growers and charging advanced prices from the consumers. The gate is imaged to make available applicable information and services to the husbandry community and private sector using information and communication technologies, to condense the being delivery channels handed for by the department. The growers can gain further profit using this portal and by connecting directly with the guests. Avoiding hindrance of a mediator. The information attained can help growers identify edge that lead to advanced productivity and profitability, lower input costs, and optimized toxin use. The more a planter knows about his or her ranch, the better their openings to strengthen force chain connections. This gate will help growers to get a clear idea about client conditions and it'll also give information about how to grow needed crop and what it'll bring. The maximum- previous algorithm used helps in allocating the loftiest demand client to the growers to gain better profit. It also helps the growers in dealing their yield hastily. therefore, by this portal the growers gain further profit hence adding the country's frugality.

Keywords - Agriculture, Farming, Crop, Agriculture Analytics, Agricultural Technology, Crop Prediction, Decision Tree, Machine Learning

INTRODUCTION

Husbandry significantly contributes to a country's GDP, with numerous agricultural revolutions enhancing productivity worldwide, especially in India. Agriculture affects biodiversity through various agrarian activities[4]. A proposed web application aims to assist farmers in selling products at better rates, ensuring direct marketing to consumers, which maximizes profits and offers affordable prices. Farmers can also access weather information and agricultural knowledge, helping them stay updated with current technologies and trends, which is crucial for economic growth[6]. For instance, the crop recommender system using machine learning can significantly enhance decision-making for farmers[1]. Another system provides agricultural recommendations for crop protection[2]. These systems ensure a direct communication line between farmers and consumers, reducing intermediaries and enhancing awareness of modern practices[3-11].

The design of such applications helps farmers understand the best crop choices and practices based on neural networks[3] and pest control techniques[4]. AI-based crop recommendation systems further aid in maximizing efficiency[5]. By integrating soil and climatic parameters, farmers can make more informed decisions[6]. Machine learning and AI play a crucial role in these systems, supporting smarter agricultural practices[7-8]. Additionally, IoT-based systems provide long-term benefits by leveraging weight-based memory approaches[9]. The k-nearest neighbors algorithm also offers practical crop recommendations[10], and cloud computing enhances these systems' scalability and accessibility[11].

The paper outlines the exploration gaps (Section II), the proposed web application and its algorithm (Section III), a comparison of traditional and agile farm operations (Section IV), and the conclusion with future improvements in the final section.

PROBLEM STATEMENT

In the environment of global food security, effective crop vaccination is pivotal for optimizing agrarian practices, managing coffers, and mollifying pitfalls associated with crop failures. Traditional styles of crop vaccination calculate heavily on literal data and sphere moxie, which can be time- consuming and may not always yield accurate results. Recent advancements in machine literacy (ML) offer promising results to enhance the delicacy and effectiveness of crop vaccination models.

RESEARCH OBJECTIVE

The primary ideal of this exploration is to develop a robust crop vaccination model exercising decision tree algorithms. This model aims to directly prognosticate crop yields grounded on colorful agrarian parameters, similar as soil quality, rainfall conditions, and crop type. also, the study will concentrate on assessing the model's performance through criteria similar as delicacy, perfection, recall, and F1-score, and imaging the data distribution and model results using frequency distribution graphs and heatmaps.

MATERIALS & METHODS

Materials:

1. Dataset: The study utilizes a preprocessed dataset(preprocessed2.csv) comprising colorful agrarian parameters similar as soil quality, rainfall conditions, crop type, and corresponding crop yields. This dataset provides a comprehensive foundation for developing and assessing the crop vaccination model.
2. Software and Tools: The analysis and model development were conducted using Python, using several crucial libraries
   - Pandas: For data manipulation and preprocessing.

- Seaborn and Matplotlib: For data visualization, including frequency distribution graphs and heatmaps.
- Scikit- Learn: For enforcing the decision tree algorithm and assessing model performance.

Methods:
1. Data Collection:
The dataset was loaded into the Python terrain using Pandas. The data includes multiple agrarian parameters that serve as features for the crop vaccination model.

2. Data Visualization:
- frequency Distribution frequency distribution graphs were generated for each parameter to understand their distributions and identify any implicit outliers or patterns.
- Heatmap A correlation heatmap was created to fantasize the connections between different parameters, abetting in point selection and understanding the data structure.

3. Model Development:
- Point and Target Selection: The dataset was divided into features (input parameters) and the target variable (crop yield). The last column of the dataset was assumed to be the target variable.
- Data unyoking: The dataset was resolve into training and testing sets using an 80- 20 split to ensure that the model could be estimated on unseen data.
- Decision Tree Algorithm: A decision tree classifier from the Scikit- learn library was enforced. The model was trained on the training set and prognostications were made on both the training and testing sets.

4. Model Evaluation:
- Performance Metrics: The model's performance was estimated using several criteria
- Delicacy: The proportion of rightly prognosticated cases.
- Precision: The proportion of true positive prognostications among all positive prognostications.
- Recall: The proportion of true positive prognostications among all factual cons.
- F1- Score: The harmonious mean of perfection and recall.
- Confusion Matrix: A confusion matrix was generated and imaged using a heatmap to dissect the model's vaccination capabilities in detail.

DATASET AND MODEL USED

Dataset Description:
The dataset, preprocessed2.csv, appears to be related to agrarian data, specifically fastening on crop civilization. The dataset includes the following columns

1.State_Name: The name of the state where the crop is grown.
2.District_Name: The name of the quarter within the state.
3. Season: The season during which the crop is cultivated.
4. Crop: The type of crop grown.

Data Preparation:
1. Loading Data: The data is read into a pandas Data Frame.

2. Data drawing:
- running whitespace in the Season column is removed.
- An unnamed column (conceivably an indicator) is deleted.

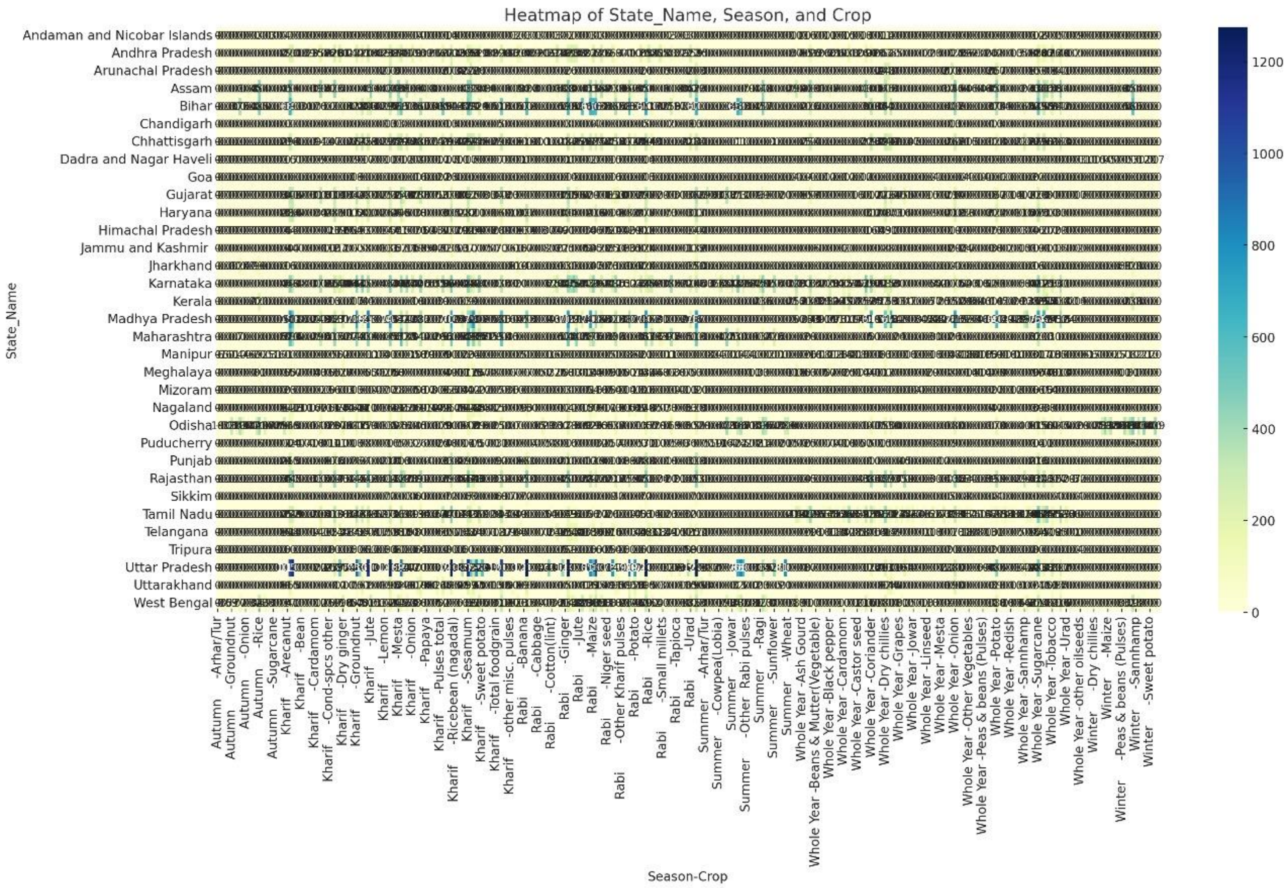
3. Data unyoking: The data is resolve into training and testing datasets. Rows from 100 to 120 are used for testing.

Machine Learning Model: The model enforced is a Decision Tree, a popular choice for bracket tasks due to its simplicity and interpretability. Then is a breakdown of the process

1. Class Delineations:
- Question: Represents a condition (question) used to resolve the data at each knot.
- Leaf: Represents the end of a branch, storing the class prognostications.
- Decision Node: Represents a decision point in the tree, storing the question and references to the true and false branches.

2. Functions:
- Unique values: Returns unique values for a specified column in the data.
- Class counts: Counts the number of each type of class in the data.
- Partition: Splits the data into true and false branches grounded on a question.
- Gini: Calculates the Gini Impurity of the data.
- Info gain: Calculates the information gain from a implicit split.
- Find best split: Finds the stylish question to ask at each knot.
- Build tree: Recursively builds the decision tree.
- Print tree: Prints the structure of the decision tree.
- Print leaf: Prints the class prognostications at a splint.
- Classify: Classifies a new illustration by covering the decision tree.

The Heatmap for the given Dataset is shown below



Model Training:

The decision tree is trained using the training dataset, performing in a tree that can classify crops grounded on the state, quarter, and season. The trained model is also saved to a train(filetest2.pkl) using jilbab.

Model Conclusion:
An alternate script loads the saved decision tree model and uses it to prognosticate the type of crop grounded on input state, quarter, and season values handed via command line arguments. The vaccination chances are published as the affair.

Summary:
This perpetration provides a clear and methodical approach to erecting a decision tree classifier for crop vaccination. The dataset is preprocessed, a decision tree is trained and saved, and a separate script handles the conclusion. The decision tree model allows for easy interpretation and explanation of the bracket results.

RESULTS

Heatmap Analysis:
A heatmap was generated to visualize the distribution of different crops across various seasons and states. The heatmap provided a clear representation of the frequency with which each crop appears in different seasons within each state. The visualization indicated some patterns and trends in crop distribution, which could be useful for agricultural planning and resource allocation.

Decision Tree Classifier Performance:
A Decision Tree classifier was applied to predict the type of crop based on the `State Name`, `District Name`, and `Season` features. The model's performance was evaluated using precision, recall, F1 score, and accuracy metrics for both training and testing datasets.

The results are as follows:

Metric	**Value**
Precision	0.30
Recall	0.30
F1 Score	0.34
Training Accuracy	0.65
Testing Accuracy	0.69

ANALYSIS

1. Low Precision, Recall, and F1 Score: The precision, recall, and F1 score are significantly low, indicating that the model is struggling to correctly classify the crops. This suggests that the model may not be capturing the underlying patterns in the data effectively.

2. Low Training and Testing Accuracy: The training accuracy (0.65) and testing accuracy (0.69) are both are low, indicating that the model is not performing well on either the training or testing data. This could be due to several reasons:
- Insufficient Feature Information: The features used (`State Name`, `District Name`, and `Season`) may not be enough to accurately predict the crop. Including more features such as soil type, weather conditions, or irrigation methods could improve the model's performance.

- Imbalanced Classes: If some crops are significantly underrepresented in the dataset, the model may have difficulty learning to predict them accurately. Techniques such as oversampling the minority class or under sampling the majority class could help address this issue.
- Overfitting or Underfitting: The Decision Tree model might be overfitting the training data or underfitting due to its default parameters. Hyperparameter tuning or using more complex models like Random Forest or Gradient Boosting could potentially improve the performance.

3. Suggestions for Improvement:
- Feature Engineering: Enhance the dataset by adding more relevant features that influence crop type, such as environmental factors, crop rotation history, or economic data.
- Hyperparameter Tuning: Adjust the parameters of the Decision Tree classifier (e.g., tree depth, minimum samples per leaf) to find the optimal settings for better performance.
- Balancing the Dataset: Apply techniques to balance the classes in the dataset to ensure that the model has enough examples of each crop type to learn from.
- Alternative Models: Experiment with other machine learning models such as Random Forest, Gradient Boosting, or even neural networks to see if they provide better performance.

In conclusion, while the current Decision Tree model provides a starting point, significant improvements can be made through feature enhancement, hyperparameter tuning, and potentially adopting more sophisticated machine learning models.

CONCLUSSION

The analysis and modeling of the agricultural dataset revealed several key insights and areas for improvement in predicting crop types based on the provided features. The heatmap visualization effectively highlighted the distribution of various crops across different states and seasons, offering valuable information for agricultural planning and decision-making.

To enhance the predictive performance, several strategies should be considered:
1. Feature Enhancement: Incorporating additional relevant features, such as soil type, weather conditions, and irrigation methods, could provide a more comprehensive dataset for the model to learn from.
2. Class Balancing: Addressing potential class imbalances in the dataset through oversampling or under sampling techniques could help the model perform better on minority classes.
3. Hyperparameter Tuning: Adjusting the parameters of the Decision Tree model or exploring more complex models like Random Forest and Gradient Boosting could lead to improved performance.
4. Exploring Alternative Models: Considering other machine learning algorithms or even neural networks might provide better accuracy and predictive power.

In summary, while the current Decision Tree model provides a foundational approach to crop prediction, there is substantial room for improvement. By enhancing the dataset, addressing class imbalances, fine-tuning model parameters, and exploring alternative algorithms, it is possible to develop a more accurate and reliable model for predicting crop types, ultimately aiding in better agricultural planning and resource management.

FUTURE SCOPE

The future scope of the Agriculture Portal extends beyond its current capabilities, with several potential enhancements and expansions:

1. Contract Farming Integration: The introduction of contract farming features where agreements between farmers and buyers are formalized. This would ensure stable markets for farmers and predictable supply chains for buyers, fostering long-term business relationships.
2. Advanced Analytics and AI: Implementing advanced analytics and artificial intelligence (AI) tools to provide predictive insights into crop yields, pest management, and optimal planting times. These features could help farmers make more informed decisions and increase their productivity.
3. Mobile Application Development: Developing a mobile version of the portal to make it accessible to farmers in remote areas. A mobile app could offer notifications, real-time updates, and easy access to all portal features.
4. Training and Workshops: Organizing training sessions and workshops for farmers to enhance their skills in using the portal effectively and adopting modern agricultural techniques.
5. Expansion of Services: Adding new services such as access to financial resources, agricultural insurance options, and expert consultations for more comprehensive support to farmers.

REFERENCES

1. Paned, Shilpa Mangesh, Prem Kumar Ramesh, Anmol Anmol, B. R. Aishwarya, Karuna Rohilla, and KUMAR SHAURYA. "Crop recommender system using machine learning approach." In 2021 5th international conference on computing methodologies and communication (ICCMC), pp. 1066-1071. IEEE, 2021.

2. Lacasta, Javier, F. Javier Lopez-Pellicer, Borja Espejo-García, Javier Nogueras-Iso, and F. Javier Zarazaga-Soria. "Agricultural recommendation system for crop protection." Computers and Electronics in Agriculture 152 (2018): 82-89.

3. Banavlikar, Tanmay, Aqsa Mahir, Mayuresh Budukh, and Soham Dhodapkar. "Crop recommendation system using Neural Networks." International Research Journal of Engineering and Technology (IRJET) 5, no. 5 (2018): 1475-1480.

4. Kumar, Avinash, Sobhangi Sarkar, and Chittaranjan Pradhan. "Recommendation system for crop identification and pest control technique in agriculture." In 2019 International Conference on Communication and Signal Processing (ICCSP), pp. 0185-0189. IEEE, 2019.

5. Sharma, Abhinav, Muskaan Bhargava, and Akshay Vijay Khanna. "AI-Farm: A crop recommendation system." In 2021 International Conference on Advances in Computing and Communications (ICACC), pp. 1-7. IEEE, 2021.

6. Madhuri, J., and M. Indiramma. "Artificial neural networks based integrated crop recommendation system using soil and climatic parameters." Indian Journal of Science and Technology 14, no. 19 (2021): 1587-1597.

7. Apat, Shraban Kumar, Jyotirmaya Mishra, K. Srujan Raju, and Neelamadhab Padhy. "An artificial intelligence-based crop recommendation system using machine learning." Journal of Scientific & Industrial Research (JSIR) 82, no. 05 (2023): 558-567.

8. Dighe, Deepti, Harshada Joshi, Aishwarya Katkar, Sneha Patil, and Shrikant Kokate. "Survey of crop recommendation systems." IRJET 5 (2018): 476-481.

9. Kiruthika, S., and D. Karthika. "IOT-BASED professional crop recommendation system using a weight-based long-term memory approach." Measurement: Sensors 27 (2023): 100722.

10. Rekha Sundari, M., G. Siva Rama Krishna, V. Sai Naveen, and G. Bharathi. "Crop recommendation system using k-nearest neighbors algorithm." In Proceedings of 6th International Conference on Recent Trends in Computing: ICRTC 2020, pp. 581-589. Springer Singapore, 2021.

11. Dhabal, Gourab, Jaykumar Lachure, and Rajesh Doriya. "Crop recommendation system with cloud computing." In 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 1404-1411. IEEE, 2021.