International Research Journal of Education and Technology
Peer Reviewed Journal
ISSN 2581-7795

AGRISMART: AI-BASED CROP PREDICTOR

Prof. KALAIVANI E
Assistant Professor
Computer Science and Engineering
Bannari Amman Institute of Technology
Erode

HARINI V
Student
Computer Science and Engineering
Bannari Amman Institute of Technology
Erode

ABSTRACT:

Crop prediction is crucial for optimizing agricultural practices, improving food security, and ensuring sustainable farming. This research focuses on evaluating various machine learning models for accurate crop prediction based on climatic, soil, and environmental data. Several state-of-the-art algorithms are tested, including Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, and Support Vector Machines, among others. These models are assessed using K-fold cross-validation to determine their predictive performance, with a particular focus on accuracy and reliability.

After evaluating multiple models, the top three performing ones are identified and selected for further enhancement. To improve the accuracy of crop prediction, a RNE (Random Forest, Naive Bayes, and Extra Trees) Stack Ensemble algorithm is employed, which combines the top three models to form a new, integrated predictive model. This algorithm leverages the complementary strengths of each base model, with a meta-learner to make final predictions. The integration of the models results in improved prediction accuracy compared to the individual models.

INTRODUCTION:

Precise crop forecasting is crucial for maximizing agricultural yield, improving food security, and promoting sustainable farming methods. In recent times, the farming industry has more frequently adopted machine learning methods to enhance the accuracy and dependability of crop forecasts by examining intricate datasets that encompass climate conditions, soil characteristics, and environmental elements. Successful crop forecasting allows farmers to utilize data-based decisions, which is essential when confronting issues like climate change, soil erosion, and changing market needs.

This research investigates the use of different machine learning models for predicting crops, assessing their efficiency and dependability in managing various agricultural datasets. Numerous reputable algorithms—like Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, and Support Vector Machines—are utilized and evaluated through K-fold cross-validation. This method guarantees thorough evaluation by examining every model's performance over several data divisions, emphasizing accuracy and reliability as vital indicators of predictive effectiveness.

Following an initial evaluation of each model's performance, the leading three models are chosen for additional enhancement via ensemble methods. To improve predictive accuracy, this study introduces the **RNE Stack Ensemble** technique, which combines Random Forest, Naive Bayes, and Extra Trees models in a stacking architecture. This ensemble model utilizes the unique strengths of each base model, with a meta-learner combining their predictions to reach a final decision. The stacking method offers a cohesive predictive model that achieves greater accuracy and dependability than individual models, highlighting the effectiveness of ensemble learning in enhancing crop prediction results.

By conducting thorough assessments, this study seeks to provide important perspectives on the effectiveness of machine learning ensembles in predicting crop yields, presenting a method that can enhance data-informed decision-making in agriculture and assist in tackling significant issues related to food security and sustainability.

LITERATURE SURVEY:

The expanding significance of ensemble learning strategies to improve model performance and minimize mistakes has been brought to light by recent developments in machine learning. Because of their ease of use and interpretability, classic models like linear discriminant analysis (LDA) and logistic regression (LR) have long served as the basis for classification tasks. Popular options for nonlinear classification include K-Nearest Neighbors (KNN) and Support Vector Machines (SVM), which are frequently used in fields needing complicated decision boundaries. Combining the advantages of many learners has resulted in notable gains in predicting performance for Random Forest, Gradient Boosting Machine, and Bagging Classifier. Because RF can handle high-dimensional data while limiting overfitting, it has drawn a lot of interest. Meanwhile, the refining of weak classifiers into stronger ones has been investigated using boosting algorithms, most notably AdaBoost and GBM.

The Stacking Classifier is an effective ensemble method that exhibits higher performance in a variety of domains by utilizing a meta-learner to combine the prediction capacity of numerous models. This study's technique is based on the recent research, which indicates that stacking can successfully overcome the constraints of individual models by merging their predictions and producing a more robust and accurate classifier.

DATASET:

There are 2,200 entries in the dataset utilized for crop prediction, and each record includes information on important agricultural characteristics. The following characteristics are present in the dataset:

- **N**: Amount of Nitrogen in the soil.
- **P**: Amount of Phosphorus in the soil.
- **K**: Amount of Potassium in the soil.
- **Temperature**: Average temperature in degrees Celsius.
- **Humidity**: Relative humidity percentage.
- **pH**: Acidity or alkalinity level of the soil.
- **Rainfall**: Average rainfall in mm.
- **Label**: Crop suitable for the given conditions.

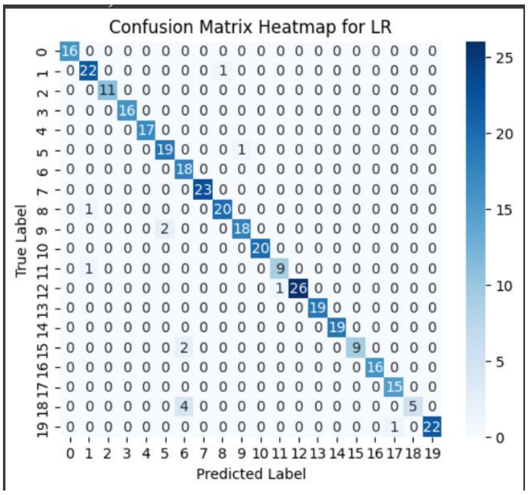
The final dataset form, following cleaning and removal of outliers, is (1768, 8). There are no duplicate values in the dataset, providing clean data for assessment and training.

PROPOSED SOLUTION:

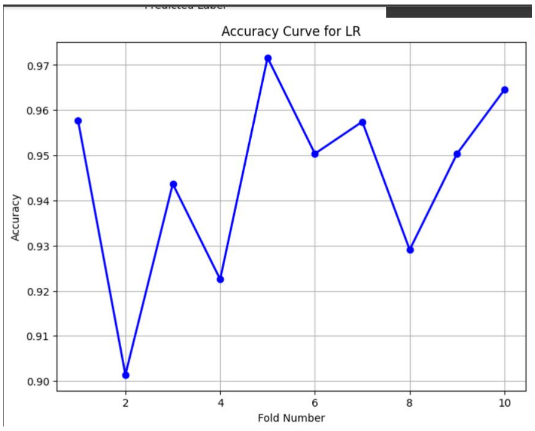
STATE-OF-ART MODELS:

Several machine learning models were tested on the dataset, and their predictive accuracies were evaluated. Below is a brief overview of the models:

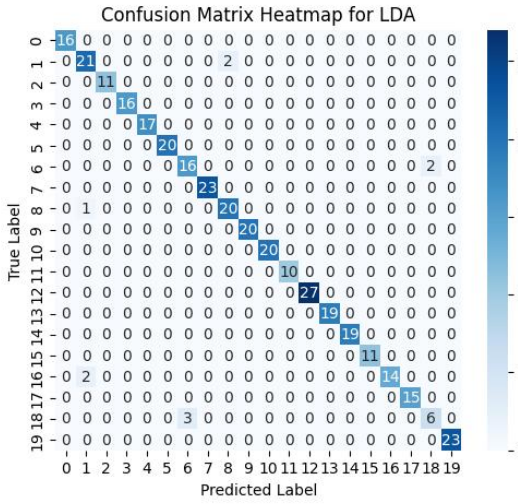
1. Logistic Regression (LR): A linear classifier frequently used for binary and multiclass classification is the logistic regression (LR) model. In order to divide the classes, it determines the optimal linear decision boundary. In our investigation, 94.49% accuracy was attained.



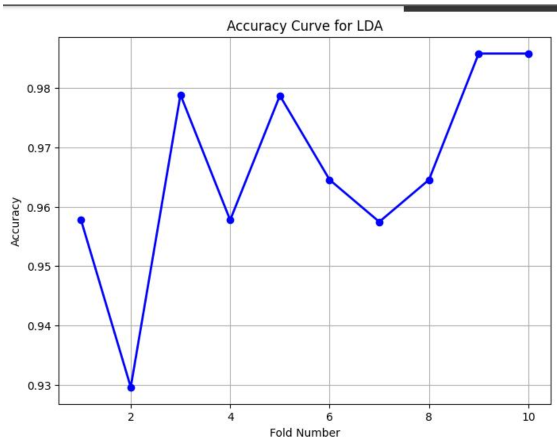
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	22	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
2	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	19	0	0	0	1	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0
8	0	1	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	2	0	0	0	18	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0
11	0	1	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	1	26	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0
15	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	9	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	0
18	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	5	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	22



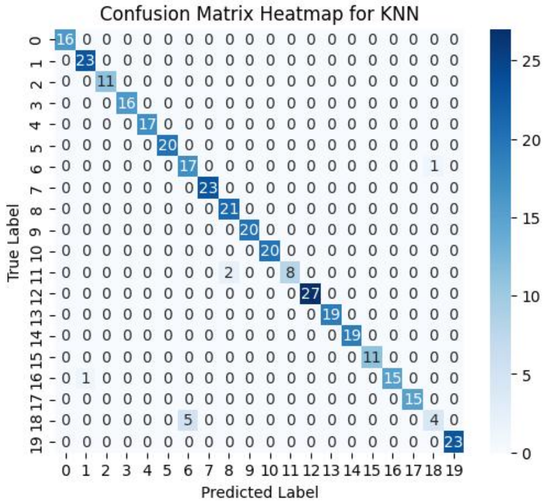
2. Linear Discriminant Analysis (LDA): By optimizing class separability, Linear Discriminant Analysis (LDA) is used to represent variations across classes. It functions effectively for dimensionality reduction and is predicated on regularly distributed classes. 96.61% accuracy was attained.

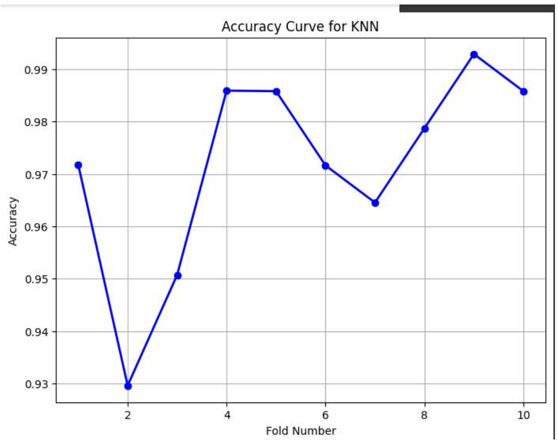


	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	21	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
2	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	16	0	0	0	0	0	0	0	0	0	0	0	2	0
7	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0
8	0	1	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0
16	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	14	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	0
18	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	6	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23

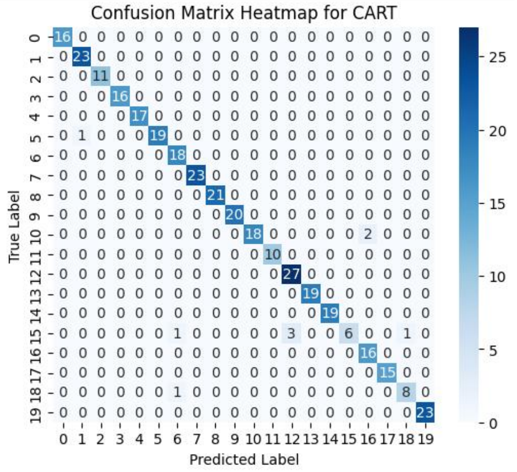


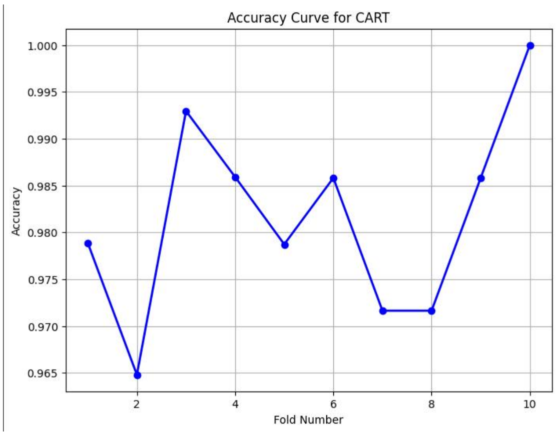
3. K-Nearest Neighbor's (KNN): The proximity to the nearest training instances is the basis for this instance-based learning technique. An accuracy of 97.17% was attained in the experimental determination of the optimal value of 'k'.



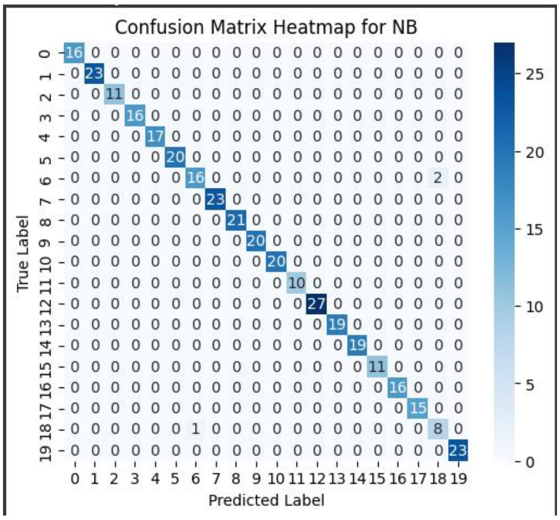


4. Classification and Regression Tree (CART): CART is a method for decision trees that divides data according to predetermined criteria in order to create a tree structure. It succeeded in obtaining a 98.16% accuracy rate.

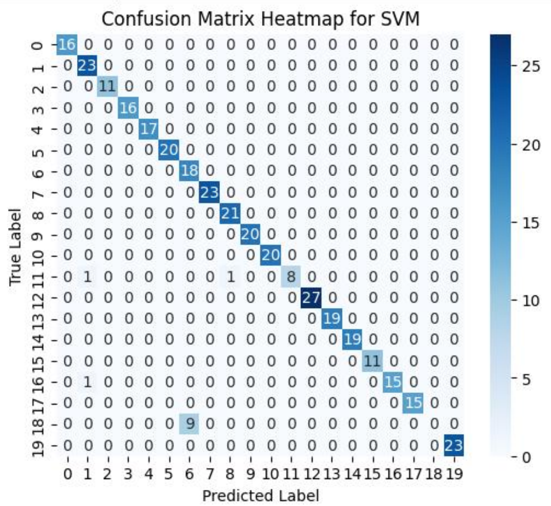




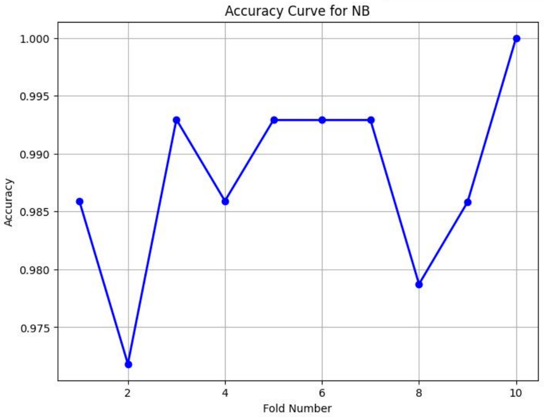
5.Naive Bayes (NB): This probabilistic classifier assumes feature independence and is based on the Bayes theorem. With an accuracy of 98.80%, it did remarkably well in spite of its simplicity.

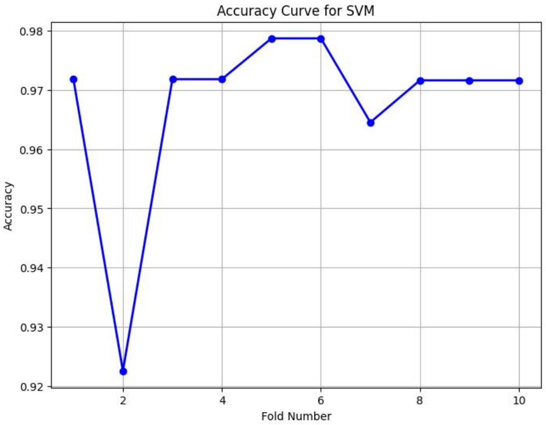


True Label	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	16	0	0	0	0	0	0	0	0	0	0	0	2	0
7	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	21	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	0
18	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	8	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23



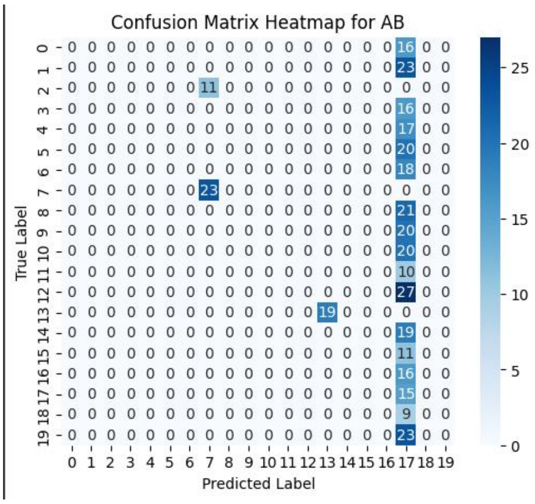
True Label	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	18	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	21	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0
11	0	1	0	0	0	0	0	0	1	0	0	8	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0
16	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	0
18	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23



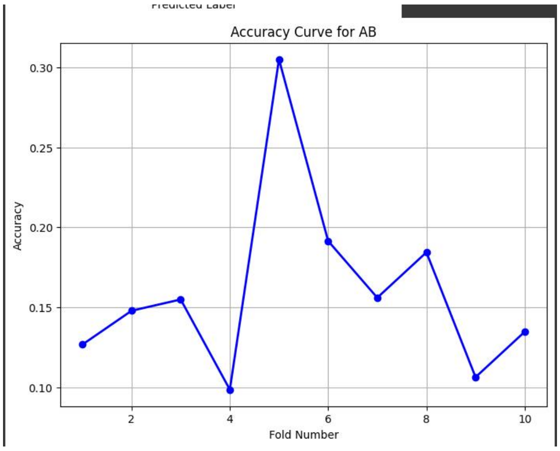


6.Support Vector Machine (SVM): SVM efficiently classifies data points by locating a hyperplane in the feature space. The model demonstrated its capacity to handle non-linear decision boundaries with an accuracy of 96.75%.

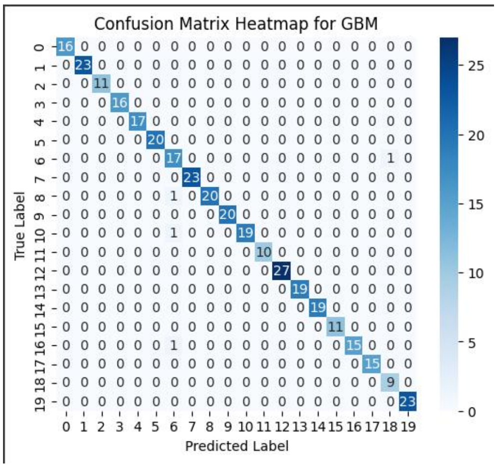
7.AdaBoost (AB): This ensemble technique creates a powerful classifier by combining several poor ones. With this dataset, though, it had difficulty; its accuracy was just 16.06%. This implies that certain weak learners were unable to adjust to the new data structure.



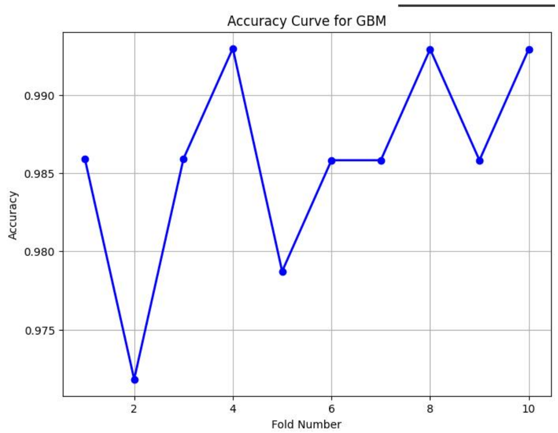
True Label \ Predicted Label	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	0	0
2	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18	0	0
7	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	0	0



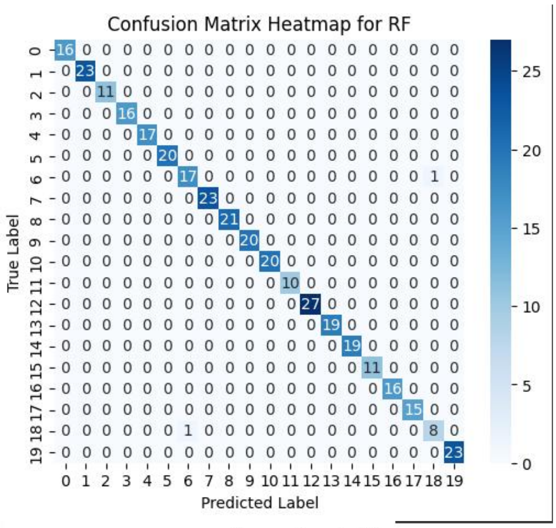
8.Gradient Boosting Machine (GBM): GBM attained an accuracy of 98.59% by combining decision trees to reduce loss repeatedly. Its capacity to recognize intricate patterns in the data contributed to its strong performance.

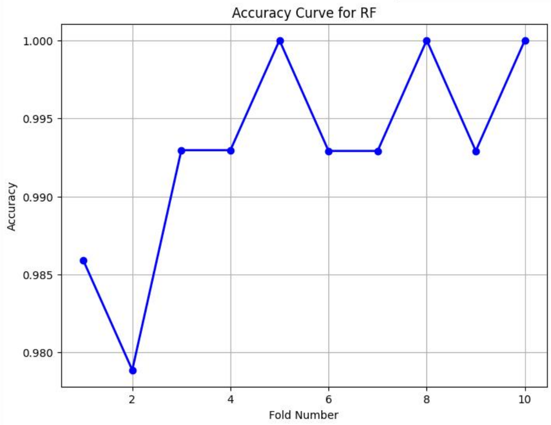


True Label \ Predicted Label	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	17	0	0	0	0	0	0	0	0	0	0	0	1	0
7	0	0	0	0	0	0	0	23	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	1	0	20	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	1	0	0	0	19	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	19	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0	0
16	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	15	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	15	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	9	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23

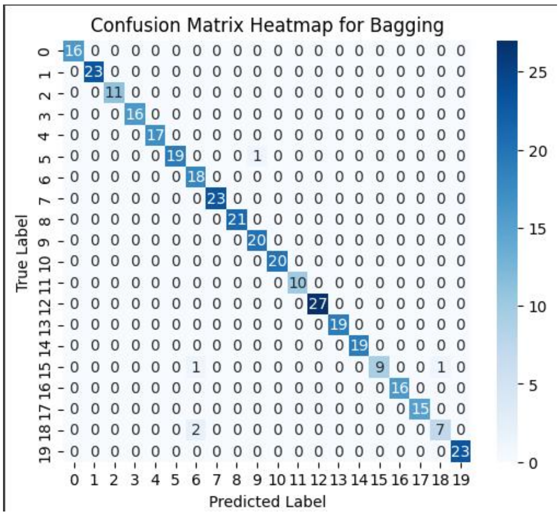


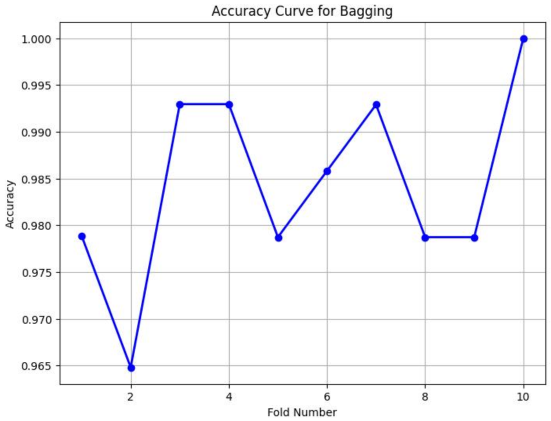
9.Random Forest (RF): RF is a collection of decision trees that functions by selecting features and data at random. It is renowned for its great precision, resilience, and skill in overcoming overfitting. With an accuracy of 99.29% in our instance, it is the highest performing single model.



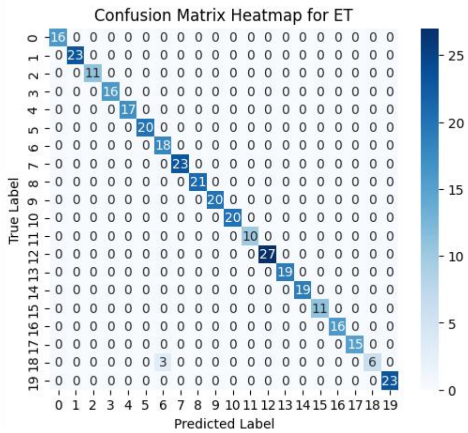


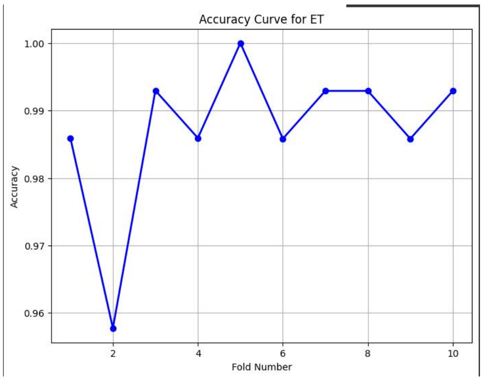
10.Bagging Classifier: Bagging is an ensemble learning method that aggregates the predictions of many decision trees. It is also known as Bootstrap Aggregation. In our analysis, it produced an accuracy of 98.44%.





11.Extra Trees (ET): Unlike RF, ET adds more unpredictability by choosing cut spots at random. This model's accuracy of 98.73% produced encouraging results.



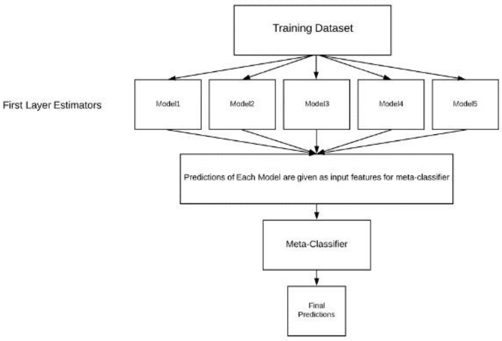


RNE STACK ENSEMBLE ALGORITHM:

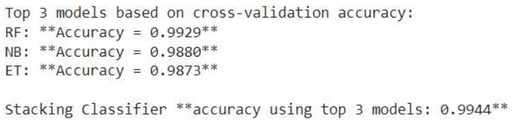
RNE (Random Forest, Naive Bayes, and Extra Trees) Stack Ensemble is a sophisticated ensemble algorithm uniquely crafted for resilient classification challenges. It utilizes a stacking classifier method that merges the advantages of Random Forest, Naive Bayes, and Extra Trees models. By combining these varied, high-performing algorithms, the RNE Stack Ensemble attains enhanced precision, robustness, and versatility across numerous data situations. This combined model is designed to enhance predictive capability by integrating probabilistic and tree-based techniques, making it an effective solution for intricate classification problems.

Stacking-classifier:

A stacking classifier is an ensemble learning method that integrates various base models, referred to as level-0 models or base learners, to form a singular, more robust predictive model. The stacking classifier is distinct from other ensemble techniques such as bagging and boosting because it incorporates an extra model, known as a meta-learner or level-1 model, above the base learners. This meta-learner uses the outputs (predictions) from the base models as its input and learns to generate final predictions from these inputs.



The concept is that every base model identifies various patterns or features of the data, and the meta-learner, usually a more straightforward model such as logistic regression, merges their predictions to produce a model that is more precise and adaptable.



```
Top 3 models based on cross-validation accuracy:
RF: **Accuracy = 0.9929**
NB: **Accuracy = 0.9880**
ET: **Accuracy = 0.9873**

Stacking Classifier **accuracy using top 3 models: 0.9944**
```

Model Selection: Using cross-validation, determined the three models with the highest accuracies:

Random Forest (RF): A strong

ensemble technique that employs various decision trees, each developed on a portion of the data, to decrease variance and enhance stability. Precision: 99.15%

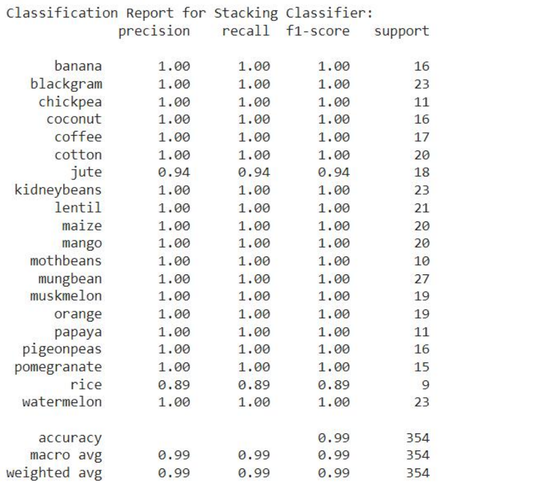
Extra Trees (ET): A different tree-based ensemble akin to Random Forest but utilizes random splits, resulting in quicker performance and less overfitting. Precision: 98.87%

Naive Bayes (NB): A statistical model derived from Bayes' theorem, recognized for its straightforwardness and efficiency with categorical data. Precision: 98.80%

Stacking Configuration: In stacking, the chosen base models (RF, ET, NB) function as level-0 models. These models undergo independent training on the training dataset, and their predictions are gathered.

Meta-Learner Training: The outputs from RF, ET, and NB are subsequently utilized as inputs for the meta-learner. In this context, we employ logistic regression as the meta-learner due to its simplicity, interpretability, and efficiency in integrating various inputs.

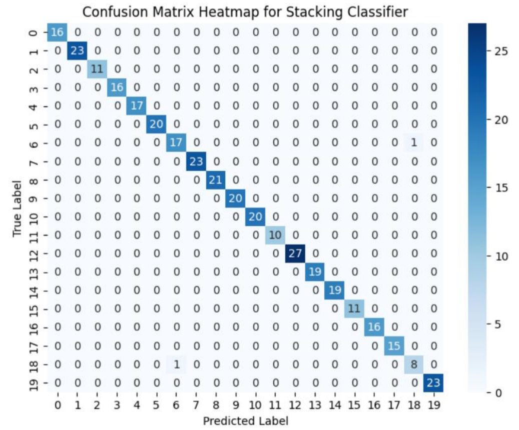
Model Assessment: Following training, the RNE Stack Ensemble is assessed using the test dataset. It attained a remarkable accuracy of 99.44%, exceeding every single model, illustrating the effectiveness of stacking.

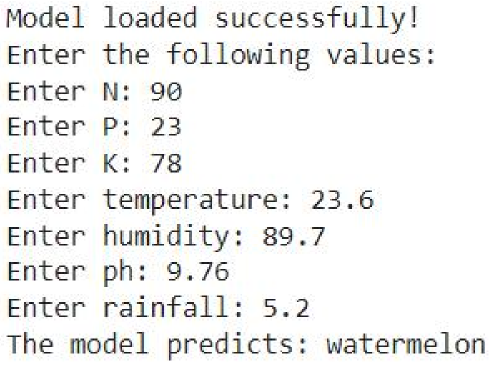


```
Classification Report for Stacking Classifier:
              precision    recall  f1-score   support

      banana       1.00      1.00      1.00        16
   blackgram       1.00      1.00      1.00        23
    chickpea       1.00      1.00      1.00        11
     coconut       1.00      1.00      1.00        16
      coffee       1.00      1.00      1.00        17
      cotton       1.00      1.00      1.00        20
        jute       0.94      0.94      0.94        18
 kidneybeans       1.00      1.00      1.00        23
      lentil       1.00      1.00      1.00        21
       maize       1.00      1.00      1.00        20
       mango       1.00      1.00      1.00        20
   mothbeans       1.00      1.00      1.00        10
    mungbean       1.00      1.00      1.00        27
   muskmelon       1.00      1.00      1.00        19
      orange       1.00      1.00      1.00        19
      papaya       1.00      1.00      1.00        11
  pigeonpeas       1.00      1.00      1.00        16
 pomegranate       1.00      1.00      1.00        15
        rice       0.89      0.89      0.89         9
  watermelon       1.00      1.00      1.00        23

    accuracy                           0.99       354
   macro avg       0.99      0.99      0.99       354
weighted avg       0.99      0.99      0.99       354
```





```
Model loaded successfully!
Enter the following values:
Enter N: 90
Enter P: 23
Enter K: 78
Enter temperature: 23.6
Enter humidity: 89.7
Enter ph: 9.76
Enter rainfall: 5.2
The model predicts: watermelon
```

CONCLUSION:

The tests showed that the RNE Stack Ensemble algorithm performed better than any of the individual models, with the best accuracy 99.44%. A more reliable and accurate forecast was made possible by the combination of Random Forest, Naive Bayes, and Extra Trees, demonstrating the value of ensemble learning in crop prediction. This survey demonstrates the potential benefits of precision agriculture and agricultural practice decision-making through the use of a stacking technique for more dependable and precise forecasts.

REFERENCES:

1.Rohit Kumar Rajak1, Ankit Pawar2, Mitalee Pendke3 , Pooja Shinde4, Suresh Rathod5, Avinash Devare6 , "Crop Recommendation System to Maximize Crop Yield using Machine Learning Technique", Dept. of Computer Engineering, Sinhagad Academy of Engineering, Maharashtra, India.

2. Dr. A. K. Mariappan, Ms. C. Madhumitha, Ms. P. Nishitha, Ms. S. Nivedhitha. (2020). Crop Recommendation System through Soil Analysis Using Classification in Machine Learning. *International Journal of Advanced Science and Technology*, *29*(3), 12738 - 12747.

3. Mafas Raheem," Crop Plantation Recommendation using Feature Extraction and Machine Learning Techniques". Soil Analysis Using Classification in Machine Learning. *International Journal of Advanced Science and Technology*, *29*(3), 12738 – 1274

4.Khaki, S., Wang, L. (2021). "Crop yield prediction integrating genotype and weather variables using deep learning." PLOS ONE. This study uses Long Short-Term Memory (LSTM) models with temporal attention to predict soybean yield based on 13 years of weather and genotype data. It highlights the importance of interpretability in machine learning models for agricultural applications.

5.Singh, G., et al. (2020). "Machine learning techniques for crop yield prediction." International Journal of Advanced Computer Science and Applications. This paper evaluates various algorithms like Decision Trees, Random Forests, and Neural Networks for yield prediction based on environmental and agricultural datasets.

6.Reddy, D.A., Dadore, B., Watekar, A. (2019). "Recommender systems for crop selection using ensemble methods." Current Agriculture Research Journal. This work discusses ensemble approaches like Random Forests and Bayesian classifiers to recommend crops based on soil and climate data, focusing on improving agricultural productivity in India.

7.Mondal, P.K., et al. (2018). "Prediction of crop yield using deep learning." Agricultural Informatics Journal. The study uses multivariate time series analysis combined with deep neural networks for

yield estimation, integrating soil, weather,
and market data