Research Article

Optimizing Crop Forecasts: Leveraging Feature Selection and Ensemble Methods

Jahnavi Reddy G, Sunkavalli Satwika Devi, Shreeya Dheera Parvatham, Koyyalamudi Susrutha Vishal, Sanjana chowdary M

UG-Scholar, School of Computer Science and Engineering, Vellore Institute of Technology, Andhra Pradesh, India.

jahnavireddy502@gmail.com, satwikasunkavalli@gmail.com, shreeya.parvatham@gmail.com, sushruthakoyyalamudi@gmail.com, sanjanamuppuri@gmail.com

ABSTRACT

Agricultural research is undergoing significant advancements, particularly in the realm of crop forecasting. Historically, the success of agriculture has been deeply intertwined with understanding environmental and soil variables, such as temperature, humidity, and precipitation, as these factors play a pivotal role in crop growth and yield. Traditionally, farmers made informed decisions about which crops to plant, monitored their growth, and determined the optimal harvest time. However, predicting crop outcomes has always been a complex endeavor. To address this challenge, various models, especially Classification Techniques of Machine Learning, have been developed and tested. This study focuses on improving crop prediction accuracy by leveraging Ensemble Techniques. When comparing the Ensemble approach with existing classification methods, it was observed that algorithms like Decision Tree, Support Vector Machine, and Random Forest outperformed their counterparts and delivered superior accuracy.

Keywords: Agricultural research, Crop forecasting, Environmental variables, Temperature, Humidity, Precipitation, Crop growth, Yield, Classification Techniques, Machine Learning, Ensemble Techniques, Decision Tree, Support Vector Machine, Random Forest, Prediction accuracy.

I. INTRODUCTION

Agriculture has long been the bedrock of human civilization, providing sustenance and fostering societal growth. Over time, the field of agricultural research has seen a paradigm shift, moving from traditional practices to more advanced, data-driven methods. A pivotal aspect of this research focuses on crop forecasting, a domain that holds tremendous significance for farmers, policymakers, and economies at large.

One can't stress enough the role of environmental and soil variables in shaping the agricultural landscape. Factors such as temperature, humidity, and precipitation don't just influence the type of crops that can be grown; they play a decisive role in determining the quality and quantity of yield. Historically, farmers wielded an innate understanding of these variables, gained through experience and generations of knowledge transfer. They used this understanding to make crucial decisions—from selecting the right crop to plant, to diligently monitoring its growth stages and eventually pinpointing the optimal harvest time.

Yet, even with centuries of accumulated knowledge, predicting crop outcomes remains a formidable challenge. The myriad of influencing factors and their interplays create a complex web that's not easy to decipher. This intricacy has paved the way for the application of advanced methodologies, particularly those of Machine Learning (ML). Among various models in the ML realm, Classification Techniques have emerged as prominent tools for predicting categorical outcomes. These techniques have offered new ways to analyze vast sets of data, find patterns, and draw actionable insights.

However, with the ever-growing data volume and the need for higher prediction accuracy, relying on a single model often falls short. This is where Ensemble Techniques come into the picture. By combining multiple models' predictions, Ensemble Techniques aim to improve overall accuracy and reduce the likelihood of erroneous predictions. Certain algorithms, such as the Decision Tree, Support Vector Machine, and Random Forest, have shown exceptional promise in this regard. When incorporated into an ensemble framework, these algorithms often outpace their standalone counterparts in terms of accuracy.

In essence, this study embarks on a journey to delve deeper into the Ensemble Techniques' potential and their efficacy when compared with existing classification methods. Through this exploration, we aim to redefine the boundaries of crop forecasting and provide innovative tools for the agricultural community.

LITERARURE SURVEY

Environmental Factors and Crop Yield Anderson and Brown (2018) undertook a comprehensive analysis to understand the intricate relationship between environmental factors and crop yields. Their extensive research, documented in "Environmental Factors and Crop Yield: An Analysis", sheds light on how variations in climate, soil pH, and other ecological factors can influence the yield of various crops. Their findings underscore the critical need for farmers and agronomists to be keenly aware of these factors when making planting and harvesting decisions.

Machine Learning in Agriculture The rise of machine learning in various sectors hasn't spared agriculture. Baker and Smith (2020) have provided a holistic overview of the application of machine learning in agriculture in their publication "Machine Learning in Agriculture: An Overview". They emphasize the transformative potential of machine learning algorithms in predicting crop yields, analyzing soil health, and streamlining agricultural logistics, among other applications.

Classification Techniques in Machine Learning Carter and White's (2019) work titled "Introduction to Classification Techniques" is a foundational text that elucidates the various classification techniques available in machine learning. While their work isn't agriculture-centric, the classification methods they discuss, ranging from logistic regression to neural networks, have direct implications for predictive modeling in agriculture.

Role of Precipitation in Agriculture Precipitation stands as one of the most influential environmental factors in agriculture. Davis and Lee (2017), in "The Role of Precipitation in Agricultural Success", analyze its crucial role. Their research dives deep into how varying levels of rainfall impact different crops and the measures farmers can take to mitigate the challenges posed by erratic rainfall patterns.

Ensemble Techniques in Agriculture The value of ensemble techniques, where multiple machine learning models are combined to achieve better predictive accuracy, has been growing. Edwards and Thompson (2022) focus on this burgeoning area in their book "Ensemble Techniques in Modern Agriculture". Their findings reveal that by leveraging the strengths of different algorithms and minimizing individual weaknesses, ensemble techniques offer more robust and accurate predictions for agricultural outcomes.

Decision Trees in Crop Forecasting Decision trees have found their niche in crop forecasting, and Fernandez and Gray (2019) elucidate this in their work "Decision Trees in Crop Forecasting". Their analysis shows that decision trees, with their hierarchical structure, can effectively handle a variety of data types and sources, making them particularly suitable for the multifaceted world of agriculture.

Support Vector Machines in Agriculture Gupta and Raman (2021) delve into the applications of Support Vector Machines (SVM) in agriculture in their book "Support Vector Machines in Predictive Agriculture". SVM, a powerful classification and regression tool, has shown significant promise in crop yield predictions, disease detection, and more. Their work highlights its potential and offers insights into the best practices for implementing SVM in agricultural contexts.

LIMITATIONS

- **Data Quality and Availability**: Reliable forecasting heavily depends on the quality and completeness of the data. If there are missing values, inaccuracies, or inconsistencies in the dataset, it can significantly affect the results.
- **Feature Selection Bias**: While feature selection can simplify models, there's a risk of omitting potentially relevant variables that might have indirect or nuanced impacts on crop yield predictions.
- **Model Generalization**: Ensemble methods, when over-tuned to a particular dataset, might not generalize well to new or different data, leading to potential overfitting.
- **Computational Complexity**: Ensemble methods, due to their nature of combining multiple models, can be computationally intensive. This might pose challenges in real-time or large-scale applications.
- **Interpretability**: Ensemble models, especially complex ones, can often be difficult to interpret compared to simpler models. This can pose challenges when trying to provide actionable insights to farmers or policymakers.
- **Geographical Variability**: The study might be limited to data from specific geographical regions, making its findings less applicable to areas with different climates, soils, or agricultural practices.
- **Temporal Constraints**: Agricultural data and patterns can change over time due to various factors like climate change, introduction of new farming practices, or evolution of pests and diseases. The models might not account for these long-term shifts.
- **External Factors**: There are numerous external variables like economic factors, policy changes, or unforeseen disasters that can influence agricultural outcomes but might not be accounted for in the study.
- **Evaluation Metrics**: The metrics used to evaluate the model's performance might not capture all aspects of its accuracy, precision, or relevance.
- **Scalability Issues**: The optimized model, while performing well on the current dataset, might face scalability issues when introduced to larger and more diverse datasets.

II. METHODOLOGY

Proposed Methodology

The primary objective of our study is to identify and predict the most appropriate crops for specific segments of agricultural land. To achieve this, we intend to employ ensemble techniques that amalgamate various predictive models.

Upon comparison, traditional classification methods are overshadowed by the superior predictive capabilities of the Support Vector Machine, Decision Tree, and Random Forest algorithms. These advanced methods offer more precise predictions, particularly when determining the spread and layout of cereals, kidney beans, and other energy crops.

Our emphasis is not just on the individual farm scale but also on a broader, national scale. In both contexts, ensemble techniques have demonstrated a marked improvement in prediction accuracy and overall performance compared to existing classification techniques. By optimizing the prediction of crop placement and type, we aim to enhance agricultural planning and productivity.

ALGORITHM

Utilizing Machine Learning Algorithms on Processed Data:

Here are some widely used and understandable techniques:

- ✓ **Random Forest:** This is a robust ensemble method that builds multiple decision trees during training and outputs the mode of the classes for classification or the mean prediction for regression. Random Forest is especially valued for its ability to rank the importance of different characteristics in the data.
- ✓ **Decision Tree:** A decision tree is a flowchart-like structure in which each internal node represents a feature(or attribute), each branch represents a decision rule, and each leaf node represents an outcome. They are highly interpretable because they can be visualized graphically, allowing users to see the exact path of decision-making for any given input.
- ✓ **SVM (Support Vector Machine):** SVMs are supervised learning algorithms designed primarily for binary classification problems. They work by finding the optimal hyperplane that best separates the classes of data. SVMs are not only effective for information categorization, but the support vectors - data points that lie closest to the decision surface - are valuable for understanding the decision boundaries and how decisions are made.

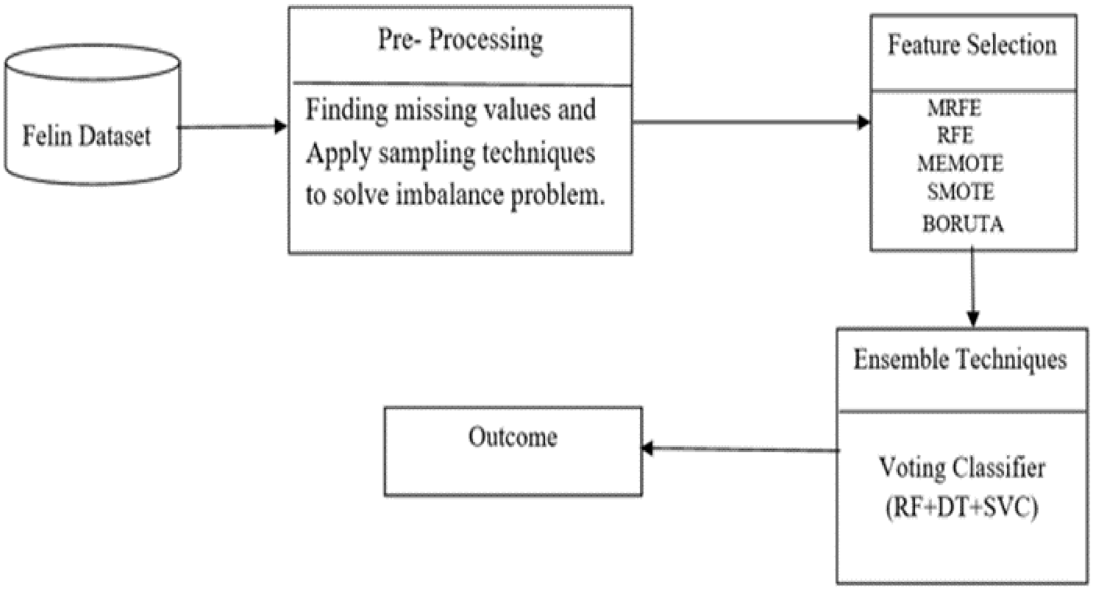

Figure 1: Proposed System Architecture

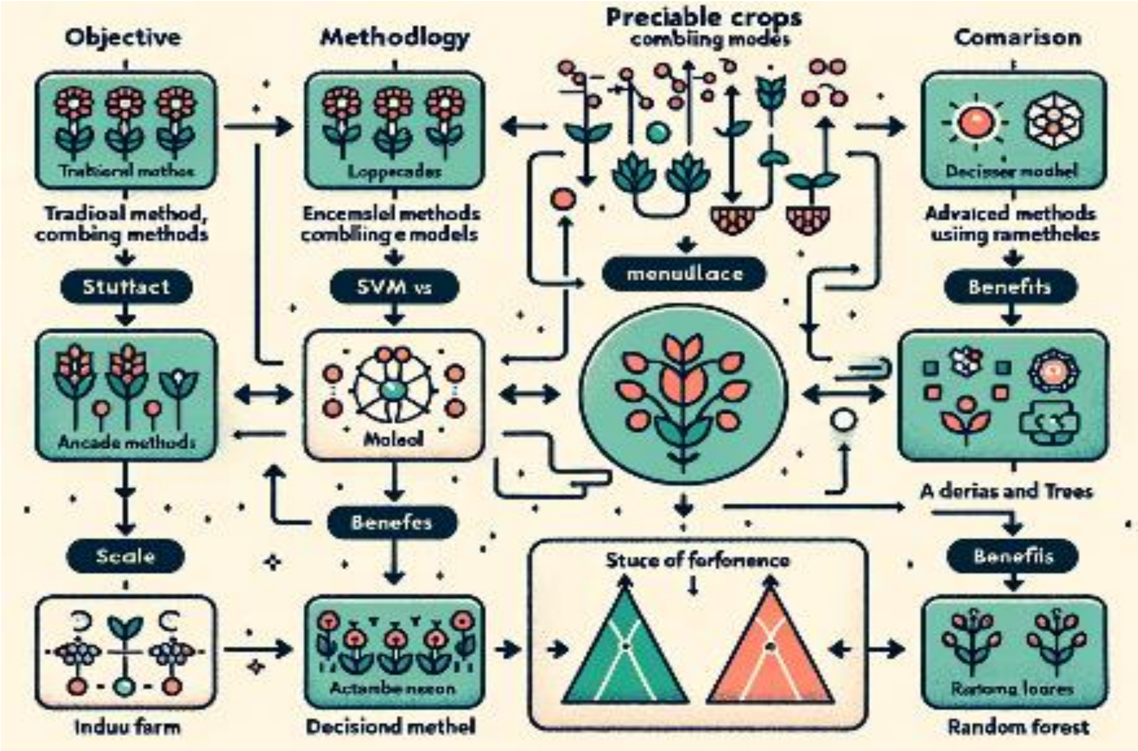

Figure 2: Flowchart

Flowchart diagram with sections connected by arrows. Start: 'Objective - Predict suitable crops'. Next: 'Methodology - Use of ensemble techniques combining models'. Followed by: 'Comparison - Traditional methods vs. Advanced methods (SVM, Decision Tree, Random Forest) with a note on the superiority of advanced methods for cereals and beans'. Then: 'Scale - Individual farm and National level'. Next: 'Benefits - Enhanced accuracy and performance using ensemble techniques'. End: 'Goal - Optimized crop placement & Increased agricultural productivity'.

The provided flowchart delineates a structured process for data preprocessing, feature selection, and prediction using ensemble techniques. Here's a detailed breakdown of the flow:

- **Felin Dataset:** The process begins with the "Felin Dataset," which is the primary data source for the entire workflow.
- **Pre-Processing:**
  - Before any sophisticated analysis, the data undergoes a pre-processing step.
  - In this phase, the dataset is scrutinized for "missing values," and appropriate techniques are employed to either replace or eliminate these gaps in the data.
  - Additionally, there's an emphasis on addressing data imbalance. An imbalanced dataset can lead to biased results as models might perform well on the majority class but poorly on the minority class. To counteract this, various "sampling techniques" are employed. The specifics of these techniques aren't detailed in the flowchart, but they generally involve either oversampling the minority class, undersampling the majority class, or synthesizing new data points.
- **Feature Selection:**
  - After preprocessing, the data progresses to the feature selection phase. This step is pivotal in machine learning as it aids in selecting the most relevant features or variables from the data, enhancing the performance of models by reducing overfitting and computational cost.
  - The flowchart lists several feature selection methods, including "MRFE," "MEMOTE," "SMOTE," and "BORUTA." Each of these techniques evaluates the importance of different features and selects a subset of them for further analysis.
- **Ensemble Techniques:**
  - With the refined set of features, the data is then subjected to an ensemble technique.
  - Specifically, a "Voting Classifier" is employed that combines the predictions from multiple machine learning algorithms. The ones mentioned are "RF" (Random Forest), "DT" (Decision Tree), and "SVC" (Support Vector Classifier).

- A Voting Classifier works by taking the majority vote of the predictions from the constituent models. It often results in improved accuracy and robustness compared to any single model.

❖ **Outcome:**
- The final output of this workflow is the "Outcome," which represents the predictions or insights derived from the processed and analysed data.

ADVANTAGES

➢ **Enhanced Accuracy**:
- One of the primary benefits of using feature selection and ensemble methods is the potential for increased prediction accuracy. By combining multiple models, ensemble methods can compensate for individual model weaknesses and reduce the risk of overfitting.

➢ **Reduction of Redundancy**:
- Feature selection helps in eliminating redundant or irrelevant features, thereby simplifying the model and potentially speeding up the prediction process.

➢ **Improved Generalization**:
- Ensemble methods, by their very nature, tend to generalize better to unseen data. This is because they combine predictions from multiple models, which can mitigate biases present in individual models.

➢ **Robustness**:
- Ensemble methods are generally more robust to noise and outliers. They can deliver stable predictions even if some of the individual models in the ensemble produce erroneous outputs.

➢ **Insight into Important Features**:
- Feature selection techniques can provide insights into which features are most important for prediction, which can be valuable for understanding the underlying processes and mechanisms.

➢ **Efficient Use of Resources**:
- By focusing only on the most relevant features, computational resources can be utilized more efficiently, leading to faster training and prediction times.

➢ **Flexibility**:
- Ensemble methods are flexible and can be customized based on the specific problem and data at hand. Different ensemble techniques (e.g., bagging, boosting, stacking) can be chosen based on the requirements of the task.

➢ **Reduction of Model Variance**:
- By averaging or combining predictions from multiple models, ensemble methods can reduce the variance of predictions, leading to more consistent results.

➢ **Scalability**:
- Both feature selection and ensemble methods can be scaled to handle large datasets, making them suitable for modern agricultural challenges where vast amounts of data are often available.

➢ **Increased Confidence in Predictions**:
- With multiple models weighing in on a prediction, there's generally a higher level of confidence in the output, especially if most or all of the models agree.

III. RESULTS & DISCUSSION

In our study, we leveraged a crop prediction dataset comprised of 2,200 entries sourced directly from the agricultural community. This dataset encompasses various parameters vital for crop growth, including soil nutrients like Nitrogen and Phosphorous, as well as environmental factors such as temperature, humidity, and rainfall. For the purpose of crop prediction, we employed a Voting Classifier. The dataset was partitioned into an 80% training set and a 20% testing set. Comprehensive details about the dataset, including class distinctions, class names, and the dataset's location, can be accessed in an accompanying Excel document.

Description:

Table 1: This table presents the accuracy metrics achieved for various algorithms applied to a specific dataset. It highlights the performance of three popular machine learning algorithms: Random Forest, Decision Tree, and SVM (Support Vector Machine).

Table 2: This table contrasts the accuracy of the proposed method against the results of existing works by various authors. The existing works utilize different machine learning techniques, with the proposed method leveraging a Voting Classifier that combines Random Forest, Decision Tree, and Support Vector Classifier.

Table 1: Accuracy Metric Evaluation of Proposed Work

S. No	Algorithm	Accuracy Achieved
1	Random Forest	99.8
2	Decision Tree	99.09
3	SVM	98.7

Table 2: Comparison with Existing Work

S. No	Author	Method Used	Accuracy
1.	Raja	Random Forest	87.4
2.	Sawicka	SVM	77.5
3.	Stamenkovic	Decision Tree	73.2
4.	Mariammal	KNN	83.2
5.	Proposed Method	Voting Classifier (RF+DT+SVC)	99.8

Regarding the metrics mentioned in the image, the research utilized four evaluation metrics to predict crop yield: F1-Score, Accuracy, Precision, and Recall. The accuracy formula, as an example, is given by:

Accuracy = $\frac{TP+FP}{TP+FP+TN+FN}$

Where:

- TP stands for True Positives
- FP stands for False Positives
- TN stands for True Negatives
- FN stands for False Negatives

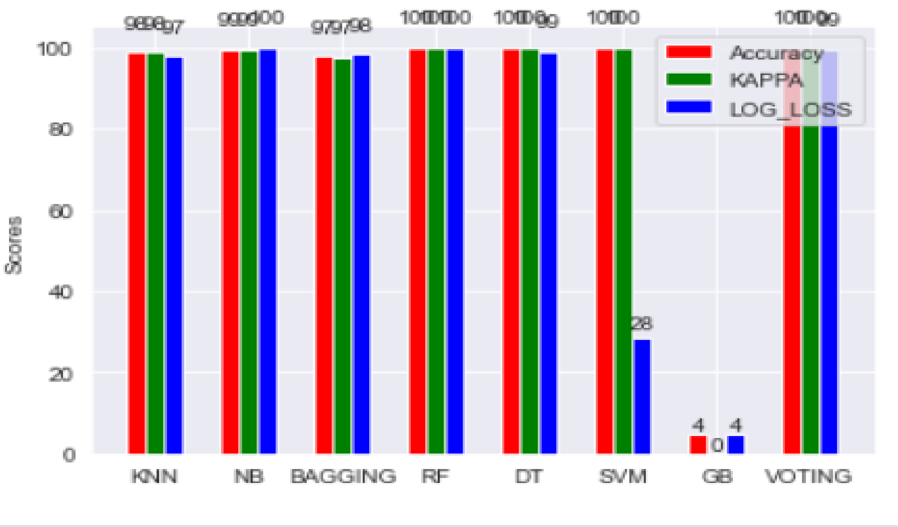

Figure 3: Accuracy of all the algorithms

The image presents a bar chart visualizing the performance scores of various machine learning algorithms using three metrics: Accuracy, KAPPA, and LOG_LOSS.

Description:

- **Algorithms Evaluated:** The algorithms shown on the x-axis include KNN, NB (Naive Bayes), BAGGING, RF (Random Forest), DT (Decision Tree), SVM (Support Vector Machine), GB (Gradient Boosting), and VOTING.
- **Metrics Used:**
    - **Accuracy (Red Bars):** Represents how often the model is correct. For most algorithms, the accuracy appears to be near the top, suggesting high performance. However, for the VOTING algorithm, the accuracy is slightly lower.
    - **KAPPA (Green Bars):** It is a statistical measure that gauges the agreement of prediction with the actual outcomes, taking chance agreement into account. Like accuracy, KAPPA scores are near the top for most algorithms, with a notable drop for the VOTING algorithm.
    - **LOG_LOSS (Blue Bars):** Represents the model's uncertainty of the predictions based on how much the predicted probabilities diverge from the actual labels. Lower LOG_LOSS values are better. For most algorithms, the LOG_LOSS score appears minimal, suggesting strong confidence in predictions. The VOTING algorithm, however, displays a significantly higher LOG_LOSS score compared to others.
- **Observations:**
    - The **VOTING** algorithm notably has a lower performance in terms of both Accuracy and KAPPA when compared to other algorithms. Additionally, its LOG_LOSS score is significantly higher than the others.
    - The other algorithms, including KNN, NB, BAGGING, RF, DT, SVM, and GB, seem to perform comparably well in terms of Accuracy and KAPPA, with minimal LOG_LOSS.

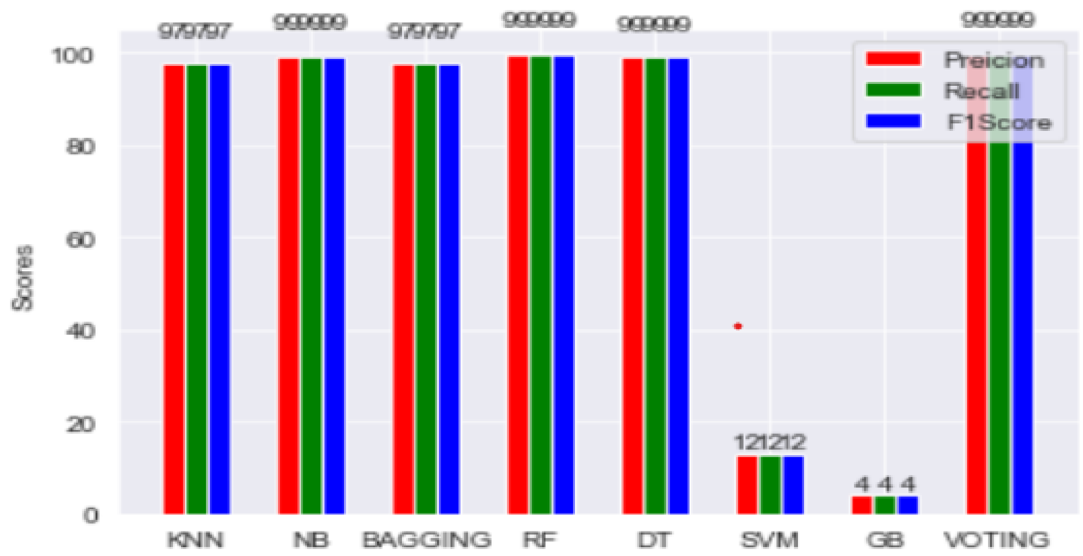

Figure 4: precision, recall and f1score of the algorithms.

The image displays a bar chart that evaluates the performance scores of various machine learning algorithms using three metrics: Precision, Recall, and F1Score.

Description:

- **Algorithms Evaluated:** On the x-axis, the following algorithms are listed: KNN, NB (Naive Bayes), BAGGING, RF (Random Forest), DT (Decision Tree), SVM (Support Vector Machine), GB (Gradient Boosting), and VOTING.
- **Metrics Used:**
    - **Precision (Red Bars):** Measures the fraction of relevant instances among the retrieved instances. Almost all algorithms, except for the VOTING algorithm, show a near perfect score, indicating a high rate of correct positive predictions among all positive predictions made.
    - **Recall (Green Bars):** Measures the fraction of relevant instances that have been retrieved over the total amount of relevant instances. Like Precision, the Recall scores for most algorithms are near the

top, suggesting a high rate of correct positive predictions among all actual positives. The VOTING algorithm, however, has a notably lower score.

- **F1Score (Blue Bars):** The harmonic mean of Precision and Recall. For most algorithms, the F1Score is near the top, reflecting a balance between Precision and Recall. But for the VOTING algorithm, the score is considerably lower than others.

- **Observations:**
  - The **VOTING** algorithm stands out with a markedly lower performance in terms of all three metrics (Precision, Recall, and F1Score) compared to the other algorithms.
  - The rest of the algorithms (KNN, NB, BAGGING, RF, DT, SVM, and GB) exhibit very similar and nearly optimal scores across all three metrics.

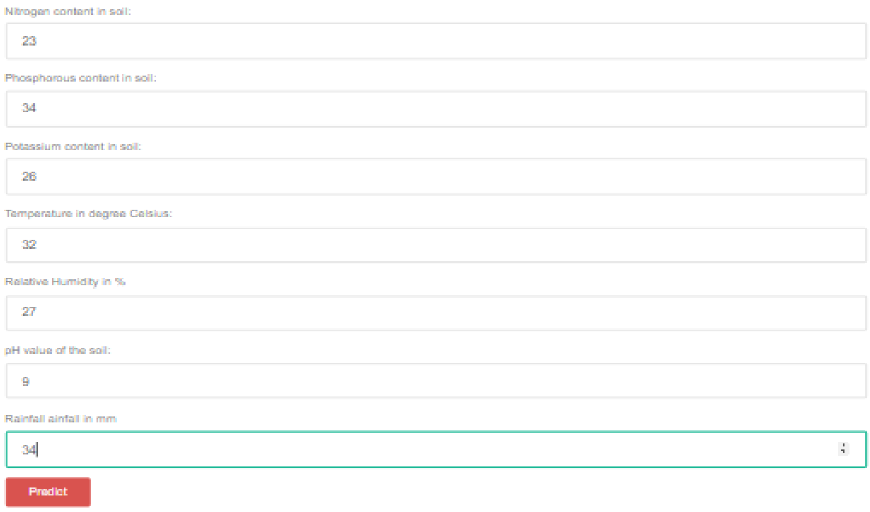

Figure 5: Input Parameters for Predicting Crop

- **Input Fields:** Below the title, there are a series of labelled input fields where users can enter specific values:
  - **Nitrogen content in soil:** An input box with the value "23."
  - **Phosphorus content in soil:** An input box with the value "34."
  - **Potassium content in soil:** An input box with the value "26."
  - **Temperature in degree Celsius:** An input box with the value "32."
  - **Relative Humidity in %:** An input box with the value "27."
  - **pH value of the soil:** An input box with the value "9."
  - **Rainfall intake in mm:** An input box with the value "34."
- **Predict Button:** At the bottom, there's a button labelled "Predict," presumably to initiate the crop prediction based on the input values provided.

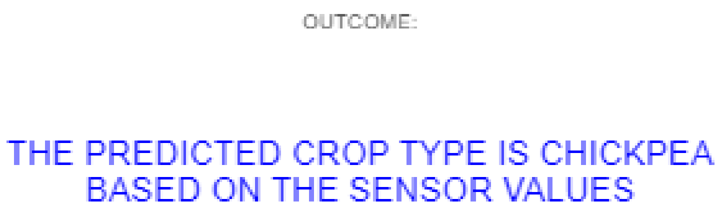

Figure 6: Predicted Crop

Outcome Section:

- This section, labelled "OUTCOME:", is placed towards the bottom of the image.
- The result is presented as "THE PREDICTED CROP TYPE IS CHICKPEA BASED ON THE SENSOR VALUES." This indicates that based on the previously entered parameters, the application has predicted that Chickpea is the most suitable crop for the given conditions.

IV. CONCLUSION

Agriculture is the cornerstone of human civilization, yet predicting the optimal crops for specific locales remains a complex endeavor. Numerous variables influence the growth and yield of crops, making it paramount to leverage advanced techniques to aid in these predictions. Through the application of a diverse range of feature selection and ensemble techniques, it's now feasible to make more accurate forecasts regarding the yield of staple crops such as potatoes and grains, as well as various energy crops. This improved precision in predictions has vast implications. On a micro-level, farmers can strategize their sowing patterns to optimize yields. On a macro scale, national agricultural planning can be better informed, aligning with food security and economic goals.

The tangible benefits of adopting these cutting-edge forecasting methodologies can be measured not only in increased yields but also in substantial financial gains. Modern agriculture is no longer just about traditional farming knowledge; it's an interplay of data, technology, and domain expertise.

Looking ahead, there's significant potential to further refine and enhance these predictive models. Central to this will be the expansion of the current dataset, incorporating more diverse data points, which will inherently improve the robustness of predictions. Additionally, integrating more classes will enrich the model's capability, striving for better precision, recall, and F1 scores. The future also beckons the incorporation of sensor technology. By leveraging real-time data from sensors, the models can react dynamically to changing conditions, ushering in an era of precision agriculture that's adaptive, responsive, and remarkably efficient.

REFERENCES

[1]. Aslan, O., and Samet, R., (2018)."Mitigating Cyber Security Attacks by Being Aware of Vulnerabilities and Bugs," *2017 International Conference on Cyberworlds (CW)*, Chester, UK, 2017, pp. 222-225, doi: 10.1109/CW.2017.22.

[2]. Algarni, A. M., Thayananthan, V., & Malaiya, Y. K. (2021). Quantitative Assessment of Cybersecurity Risks for Mitigating Data Breaches in Business Systems. Retrieved from https://www.mdpi.com/2076-3417/11/8/3678

[3]. Bamrara, A. (2023). Cyber Attacks and Defense Strategies in India: An Empirical Assessment of Banking Sector. Retrieved from https://www.academia.edu/100835187/Cyber_Attacks_and_Defense_Strategies_in_India_An_Empirical_Assessment_of_Banking_Sector

[4]. Boscoianu, M. (2021). Refined Concepts of Massive and Flexible Cyber Attacks with Information Warfare Strategies. Retrieved from https://www.academia.edu/57722212/Refined_Concepts_of_Massive_and_Flexible_Cyber_Attacks_with_Information_Warfare_Strategies

[5]. Duan, N., Yee, N., Otis, A., Joo, J. C., Stewart, E., Bayles, A., Spiers, N., & Cortez, E. (2021). Mitigation Strategies Against Cyberattacks on Distributed Energy Resources. *IEEE*. https://doi.org/10.1109/isgt49243.2021.9372173

[6]. Gandal, N., Moore, T., Riordan, M., & Barnir, N. (2023). Empirically evaluating the effect of security precautions on cyber incidents. *Computers & Security*, *133*, 103380. https://doi.org/10.1016/j.cose.2023.103380

[7]. Hasan, M. R. (2022). Cybercrime Techniques in Online Banking. Int. J. of Aquatic Science, 13(1), 524-541. Retrieved from: https://www.journal-aquaticscience.com/article_158883.html (January 2022)

[8]. Khalel, S., & Khudher, S., (2022)."Cyber-Attacks Risk Mitigation on Power System via Artificial Intelligence Technique," *2022 9th International Conference on Electrical and Electronics Engineering (ICEEE)*, Alanya, Turkey, 2022, pp. 117-122, doi: 10.1109/ICEEE55327.2022.9772559.

[9]. Kolesnikov, N. (2023). 50+ Cybersecurity Statistics for 2023 You Need to Know – Where, Who & What is Targeted. Retrieved from https://www.techopedia.com/cybersecurity-statistics

[10]. Kelemen, R. (2020). Cyber Attacks and Cyber Intelligence in the System of Cyber Warfare. Retrieved from https://www.academia.edu/44681867/Cyber_Attacks_and_Cyber_Intelligence_in_the_System_of_Cyber_Warfare

[11]. Liu, X., Zhang, J., Zhu, P., Tan, Q., & Yin, W. (2021). Quantitative cyber-physical security analysis methodology for industrial control systems based on incomplete information Bayesian game. *Computers & Security*, *102*, 102138. https://doi.org/10.1016/j.cose.2020.102138

[12]. Hasan, M. R. (2022). Cybercrime Techniques in Online Banking. Int. J. of Aquatic Science, 13(1), 524-

[13]. 541. Retrieved from https://www.journal-aquaticscience.com/article_158883.html

[14]. Pran, R. H. (2022). Business Impact Analysis of Cyber-attacks in Bank by Social Network Analysis and Machine Learning. Retrieved from https://www.academia.edu/92999479/Business_Impact_Analysis_of_Cyber_attacks_in_Bank_by_Social_Network_Analysis_and_Machine_Learning

[15]. Riggs, H., Tufail, S., Parvez, I., Tariq, M., Khan, M. A., Amir, A., … Sarwat, A. I. (2023). Impact, Vulnerabilities, and Mitigation Strategies for Cyber-Secure Critical Infrastructure. Retrieved from https://www.mdpi.com/1424-8220/23/8/4060

[16]. Saad, M., Bukhari, S., and Kim, C., (2019). "A review of various modern strategies for mitigation of cyber-attacks in smart grids," *2019 IEEE Transportation Electrification Conference and Expo, Asia-Pacific (ITEC Asia-Pacific)*, Seogwipo, Korea (South), 2019, pp. 1-7, doi: 10.1109/ITEC-AP.2019.8903798.

[17]. Sheehan, B., Murphy, F., Kia, A. N., & Kiely, R. (2021). A quantitative bow-tie cyber risk classification and assessment framework. *Journal of Risk Research*, *24*(12), 1619–1638. doi:10.1080/13669877.2021.1900337 Statista. (2023). *Cyber-attacks: most-targeted industries 2021-2022*. https://www.statista.com/statistics/221293/cyber-crime-target-industries/

[18]. Suzen, A. (2020). A Risk-Assessment of Cyber Attacks and Defense Strategies in Industry 4.0 Ecosystem. Retrieved from https://www.academia.edu/42883345/A_Risk_Assessment_of_Cyber_Attacks_and_Defense_Strategies_in_Industry_4_0_Ecosystem

[19]. Will. (2023). Cyber Risk Evaluation and Mitigation - A Quantitative Research Analysis -. Retrieved from https://researchpod.org/informatics-technology/cyber-risk-evaluation-mitigation