Introduction:

Agriculture is the foundation of human civilization, providing nutritional, livelihood and economic growth. However, as the world's population increases, the need for agricultural resources increases, necessitating a review of conventional farming practices. In order to meet the requirements for food from a growing population and boost yields, modern agriculture must also work to minimize its impact on the environment. According to the latest research, current agricultural food production must increase by more than 70% by 2050 (Thilakarathne et al., 2022).This need is driven by the need to earn a living for an expanding global population in. Maintaining this delicate balance desperately needs precision agricultural solutions that maximize resource efficiency, increase productivity, and reduce environmental impact.

The introduction of advanced technology in agriculture has led to the development of various intelligent agricultural products. Among these, the Crop Recommendation Scheme (CRS) has emerged as a promising approach to address the challenges of sustainable agricultural production CRS can use artificial intelligence technology, machine learning, and data analytics to make specific crop recommendations based on a variety of variables, including soil type, climate, historical yield data, market trends, etc., to make decisions consistent with their economic interests and environmental considerations (Jadhav & Bhaladhare, 2022).

As the world faces food security and conservation challenges, the role of technology in reshaping agriculture is becoming increasingly apparent Although the Crop Recommendation System (CRS) has made remarkable progress in changing the crop selection through data analysis -To provide synthetic yet accurate information to CRS benefits generated by agriculture and knowledge integration (Mani & Edinburgh, n.d.). CRS based on data-driven insights is only as robust as data capture. Finding the optimal mix of data types and volumes can be a challenging endeavor. Real-world agricultural datasets, especially at the granular level, are often scarce and variable in time and space (Chougule et al., 2019). This shortcoming hinders the flexibility and predictive accuracy of the system, especially in dynamic environments such as agricultural land.

The ability of Generative Adversarial Networks (GANs) to generate synthetic data that closely resembles extensive real-world segmentation has been proven (Li et al., 2019). In our approach, GANs work together with rule-based algorithms to generate synthetic farm records. GAN’s generator crafts data that match the underlying distribution of real agricultural records. A discriminator trained on real-world data ensures that the generated models are consistent with realistic, complex agricultural data.

The official website of Tamil Nadu Agriculture serves as a goldmine of valuable agricultural statistics, encompassing diverse variables such as soil composition, weather patterns, crop yields, and cultivation practices. This repository not only serves as a rich source of authentic data but also informs the rule-based algorithms guiding our GANs. By encapsulating the wisdom of Tamil Nadu's agricultural practices, we augment the realism and relevance of the synthetic data generated.

The synergy between GAN-generated synthetic data and rule-based algorithms imparts a newfound vitality to Crop Recommendation Systems. The enriched datasets overcome the limitations of real-world data scarcity, capturing relationships and subtle changes. Using this synthesis of data, CRS’s recommendations are deeper and more accurate, enabling farmers to gain insights that reflect a harmonious mix of established practices and new technologies.

The ensemble method shows a symbiotic unity between gradient boosting and SVM. Gradient boosting excels in capturing complex relationships and patterns, while SVM results in complex decision boundaries (Guerrero et al., 2012). By combining their results in a weighted manner, we develop an integrated model that provides comprehensive insights into crop selection (Fan et al., 2018).

This approach not only overcomes the challenge of data scarcity, but also promotes CRS with the predictive power of advanced machine learning. As we move forward on the journey of sustainable agriculture, this initiative will emerge as a beacon of innovation, promising to transform the way crops are grown, adapted and managed.

Literature review:

According to (Ujjainia et al., 2021), The impact of technology on agriculture cannot be ignored, as it has led many agricultural organizations and farmers to incorporate smart and innovative technologies into their practices and this trend has subsequently led to a significant increase in crop production, leading to the needs of the growing global population. However, the agricultural industry is facing an ongoing challenge to accurately predict the complexity of the ecological systems required for optimal yields and in this case, the application of machine learning appears as a potential solution to this issue (Gosai et al., 2021).

Machine learning techniques find utility in agriculture, where historical data sets are analyzed. These research methodologies have the potential to provide valuable insights for farmers and the national economy as a whole (Suresh et al., n.d.). (Madhuri & Indiramma, 2021) employed artificial neural networks to make agricultural recommendations depending on the features of the crop, the soil, and the climate. This recommendation model uses real-time input to advise the best crop for a given area. In general, there are only a few dimensions that machine learning models may handle. The model's complexity grows as the variety of variables rises. (Attaluri et al., 2020).

The reliability of these datasets directly impacts the accuracy and effectiveness of machine learning algorithms in providing informed crop suggestions. This paper introduces PSO-MDNN, recommendation model created using a classifier and an optimization of the classifier using a Modified DNN (Deep Neural Networks) and PSO (Particle Swarm Optimization) for crop cultivation (Mythili & Rangaraj, 2021). This suggests that deep learning approaches can be less expensive and scale well in locally collected samples. In this (Li et al., 2019) Generative Adversarial Networks (GANs) is employed to create new data instances that resemble the distribution of the original dataset, effectively increasing the diversity and size of the training data available for machine learning models. A generator and a discriminator comprise the initial GAN. In the original GAN, the generator's input is random noise (Tian et al., 2021).

The Support Vector Machine (SVM) is a non-parametric classification method that uses nonlinear kernel operations to distinguish between multimodal class distributions in feature spaces with high dimensions (Löw et al., 2013). It performs poorly on overlapping classes but has minimal effect on outliers (Ganesan et al., 2021)..

During training, many established AI models are intricate and computationally costly. Tree-based ensembles such as Gradient Boosting (GB), Random Forest (RF), and Extremely Randomized Trees (Extra-Trees) are examples of rule-based Decision Trees (DT) are gaining prominence due to their simplicity, strength, and predictive prowess (Fan et al., 2018). An analysis of the recommendation system describes methods of policy recommendation e.g. Collaborative, Content-Based Hybrid Recommender Systems and their limitations and possible extensions for those who can improve the recommendation capabilities of these systems (Kuanr et al., 2018).

This study (Oikonomidis et al., 2022), shows that the second best results were obtained with the XGBoost model, which took less time to implement than other DL-based models. By using efficient feature selection approaches and a heterogeneous methodology for estimating crop production, they had accounted for the majority of the necessary predictor factors (Iniyan & Jebakumar, 2022).

Methodology:

Raw data

Rule Based Data

Generative Adversial Network (GAN)

Extreme Gradient Boosting (XGBoost)

Support Vector Machine (SVM)

Ensemble Model

Data Generation:

**Rule-Based Approach:**

A rule-base­d approach has been deve­loped to generate­ synthetic agricultural data, which serves as a vital re­source for various applications in agriculture. Our data gene­ration algorithm involves several ke­y steps:

1. **Data Retrieval**: We extracted relevant attributes, including crop types, soil properties, seasonal information, and more, from an agricultural dataset.
2. **Data Augmentation**: To introduce realistic variation, we augmented the dataset by varying attributes such as soil pH, crop duration, temperature, water source, water requirements, and relative humidity.
3. **Data Simulation**: Using randomization techniques, we simulated data points based on predefined rules. For instance, soil pH values were generated within specified ranges, mimicking real-world soil conditions. Similarly, crop duration, temperature, and other attributes were simulated based on agricultural norms.
4. **Data Integration**: The generated data points were integrated into a new dataset that closely resembles real-world agricultural data.

The rule-based approach successfully generated a synthetic agricultural dataset with attributes reflecting crop characteristics, soil properties, seasonal factors, and environmental conditions. The introduced variability ensures that the synthetic dataset approximates real-world agricultural scenarios.

**Data Integration with GAN Model:**

To further enhance the utility of synthetic agricultural dataset, we integrated it into the Generative Adversarial Network (GAN) model. GANs are widely known for their ability to generate accurate data that resemble large-scale training data distributions. Our GAN model has two main features:

1. **Generator**: The generator is responsible for generating synthetic agricultural data points based on the patterns and characteristics learned from the real and augmented data. It takes random noise as input and produces synthetic data samples.
2. **Discriminator**: The discriminator evaluates the authenticity of the generated data samples. It distinguishes between real agricultural data and synthetic data produced by the generator.

We employed a training process to optimize our GAN model using the synthetic agricultural dataset:

1. **Data Preprocessing**: The synthetic dataset, which includes crop types, soil properties, and environmental factors, was preprocessed to ensure compatibility with the GAN model. Data scaling and encoding were performed as necessary.
2. **Model Training**: The generator and discriminator were trained iteratively. The generator aimed to produce synthetic data that could not be distinguished from real data by the discriminator, while the discriminator aimed to correctly classify real and synthetic data.
3. **Convergence**: The GAN model was trained until convergence, ensuring that the synthetic data generated by the generator closely matched the patterns observed in the real agricultural dataset.

The integration of the synthetic agricultural dataset with the GAN model has yielded promising results. The GAN-generated synthetic data closely resembles real-world agricultural data in terms of attributes and patterns. This approach enhances the data augmentation process by generating diverse data points that adhere to the underlying rules of agricultural systems.

Ensemble Approach:

Ensemble techniques often lead to more robust and accurate models, which is advantageous in a CRS where the accuracy of recommendations is vital. Ensemble learning combines multiple machine learning models to enhance predictive accuracy. In a Crop Recommendation System (CRS), Support Vector Machines (SVM) contribute robust, non-linear modeling, while Gradient Boosting offers powerful sequential learning. SVMs and Gradient Boosting are chosen for their individual strengths, complementarity, and ability to handle complex agricultural data, making them key components in improving crop recommendations within the ensemble approach. The combination of these models leverages diverse insights and promotes more accurate and stable crop recommendations.

Support Vector Machine:  
  
Support Vector Machinеs (SVM) arе a powerful class of supervised machine learning algorithms primarily used for classification and regression tasks. Thеy arе particularly valuablе in situations where data exhibits complex decision boundariеs. SVMs are chosen for their ability to handle complex data, prevent overfitting, adapt to various data structures through kernel functions, and provide well-defined decision boundaries. These characteristics make SVMs a valuable tool in machine learning for a wide range of applications, including agricultural tasks like crop recommendation, where the relationship between crop selection and various factors can be intricate and non-linear.

Support Vector Machines (SVM) have gained prominence in various fields due to their ability to handle both linear and non-linear classification tasks effectively. In the context of our Crop Recommendation System, several factors make SVM a compelling choice:

Non-linearity of Data: Agricultural datasets often exhibit non-linear relationships between various parameters, such as soil type, water source, and crop type. SVM can capture these intricate relationships and make accurate predictions.

Robustness: SVM is robust against overfitting, which is particularly important when dealing with agricultural data that may have inherent noise and variability.

High-Dimensional Data: Agricultural datasets often involve numerous features or parameters. SVM can handle high-dimensional data efficiently, making it suitable for our multi-feature crop recommendation task.

Flexibility: SVM provides flexibility in choosing different kernel functions, such as linear, polynomial, or radial basis function (RBF) kernels. This adaptability allows us to experiment with different models based on the nature of the data.

The dataset generated from GAN is split into training and testing sets using the train\_test\_split function. This separation ensures that the model is trained on one subset and evaluated on an independent subset to assess its generalization capabilities. The features are standardized using the StandardScaler to ensure that each feature has a mean of 0 and a standard deviation of 1. Feature scaling is essential for SVMs, as they are sensitive to feature scales. The SVM model is constructed using the SVC (Support Vector Classification) class from Scikit-Learn. It uses a linear kernel and a regularization parameter (C) set to 1.0. The SVM model is trained on the standardized training data using the fit method. Predictions are made on the standardized test data, and the model's performance is assessed using accuracy and a classification report, which includes precision, recall, F1-score, and support for each class.

Accuracy: 1.0

Classification Report:

precision recall f1-score support

0 1.00 1.00 1.00 9

1 1.00 1.00 1.00 11

2 1.00 1.00 1.00 9

3 1.00 1.00 1.00 10

4 1.00 1.00 1.00 8

5 1.00 1.00 1.00 8

6 1.00 1.00 1.00 5

7 1.00 1.00 1.00 12

8 1.00 1.00 1.00 10

9 1.00 1.00 1.00 6

10 1.00 1.00 1.00 12

11 1.00 1.00 1.00 6

12 1.00 1.00 1.00 9

13 1.00 1.00 1.00 11

14 1.00 1.00 1.00 9

15 1.00 1.00 1.00 15

16 1.00 1.00 1.00 14

17 1.00 1.00 1.00 10

18 1.00 1.00 1.00 9

19 1.00 1.00 1.00 11

20 1.00 1.00 1.00 8

...

accuracy 1.00 570

macro avg 1.00 1.00 1.00 570

weighted avg 1.00 1.00 1.00 570

Gradient Boosting:

Gradient Boosting is known for its exceptional predictive accuracy. In a Crop Recommendation System, accurately recommending the most suitable crops to farmers is critical for optimizing yields and resource utilization. The high accuracy of Gradient Boosting can lead to more precise crop recommendations. Agricultural data often involves complex, non-linear relationships between various factors such as soil type, weather conditions, and crop types. XGBoost, an extreme version of Gradient Boosting is known for its high predictive accuracy, is pivotal in Crop Recommendation Systems (CRS). It excels in capturing complex, non-linear patterns and relationships in agricultural data, enhancing precision in crop recommendations. XGBoost's ensemble approach combines weak learners, fortifying predictive accuracy. It quantifies feature importance, aiding agronomists in understanding influential factors like soil type and climate variables. XGBoost's scalability accommodates extensive, diverse agricultural datasets, optimizing resource utilization in farming

In our research, we harnessed the XGBoost classifier, a potent tool for our Crop Recommendation System. We initiated by loading our dataset and encoding categorical variables using Label Encoding to prepare it for machine learning. The dataset was split into training and testing sets (80% and 20%, respectively) for model evaluation. We leveraged XGBoost with 'multi: softmax' as the objective function to accommodate multiple crop classes. After training, we made predictions on the test set and computed the accuracy score. Additionally, we generated a classification report, highlighting precision, recall, and F1-scores for each crop class. These results offer valuable insights into the model's performance and its potential to optimize crop recommendations, which is a fundamental aspect of our research."

Accuracy: 0.9947368421052631

Classification Report:

precision recall f1-score support

0 1.00 1.00 1.00 9

1 0.92 1.00 0.96 11

2 1.00 1.00 1.00 9

3 1.00 1.00 1.00 10

4 1.00 1.00 1.00 8

5 1.00 1.00 1.00 8

6 1.00 1.00 1.00 5

7 1.00 1.00 1.00 12

8 1.00 1.00 1.00 10

9 1.00 1.00 1.00 6

10 1.00 1.00 1.00 12

11 1.00 1.00 1.00 6

12 1.00 1.00 1.00 9

13 1.00 1.00 1.00 11

14 1.00 1.00 1.00 9

15 1.00 1.00 1.00 15

16 1.00 1.00 1.00 14

17 1.00 1.00 1.00 10

18 1.00 1.00 1.00 9

19 1.00 1.00 1.00 11

20 1.00 1.00 1.00 8

21 1.00 1.00 1.00 13

...

accuracy 0.99 570

macro avg 1.00 1.00 0.99 570

weighted avg 1.00 0.99 0.99 570

The Ensemble Model using the VotingClassifier from the sklearn.ensemble module leverages the strengths of both SVM (Support Vector Machine) and Gradient Boosting to potentially enhance the overall predictive accuracy. The ensemble employs 'soft' voting, meaning it considers probability-based predictions from both SVM and Gradient Boosting. This approach considers the confidence of each model's predictions, potentially improving the overall accuracy of the ensemble.

Conclusion:

An in-depth analysis of Crop Recommendation Systems (CRS) with an emphasis on data generation and ensemble modeling. In order to generate data, we integrated rule-based methods with generative adversarial networks (GANs). Agricultural principles and domain expertise were included into the rule-based approach to lay the groundwork, and GANs contributed complexity and diversity by learning and recreating complex patterns from the actual world. In our ensemble model, the Support Vector Machine (SVM) and XGBoost delivered promising outcomes. SVM made use of its robustness in classification tasks, and XGBoost identified intricate connections in the data. By utilizing these models' synergies, the ensemble technique increased overall predictive accuracy in the CRS domain. Crop recommendations might be revolutionized by using this all-encompassing strategy, which would also maximize agricultural productivity and resource use. Deeper insights into the ensemble's efficiency in improving crop recommendations for precision agriculture will come from further examination, including precision, recall, and F1-scores.

Result:

The Ensemble model, combining the strengths of both SVM and Gradient Boosting, demonstrated a competitive performance. This approach harnesses the diversity of the individual models, potentially enhancing overall predictive accuracy in the context of your Crop Recommendation System. Study of precision, recall, and F1-scores, provides deeper insights into the ensemble's effectiveness in optimizing crop recommendations.

Ensemble Accuracy: 0.9964912280701754

References:

Attaluri, S. S., Batcha, N. K., & Mafas, R. (2020). Crop Plantation Recommendation using Feature Extraction and Machine Learning Techniques. In *Journal of Applied Technology and Innovation* (Vol. 4, Issue 4).

Chougule, A., Jha, V. K., & Mukhopadhyay, D. (2019). Crop suitability and fertilizers recommendation using data mining techniques. *Advances in Intelligent Systems and Computing*, *714*, 205–213. https://doi.org/10.1007/978-981-13-0224-4\_19

Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X., Lu, X., & Xiang, Y. (2018). Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energy Conversion and Management*, *164*, 102–111. https://doi.org/10.1016/j.enconman.2018.02.087

Ganesan, M., Andavar, S., & Raj, R. S. P. (2021). Prediction of Land Suitability for Crop Cultivation Using Classification Techniques. *Brazilian Archives of Biology and Technology*, *64*. https://doi.org/10.1590/1678-4324-2021200483

Gosai, D., Raval, C., Nayak, R., Jayswal, H., & Patel, A. (2021). Crop Recommendation System using Machine Learning. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 558–569. https://doi.org/10.32628/CSEIT2173129

Guerrero, J. M., Pajares, G., Montalvo, M., Romeo, J., & Guijarro, M. (2012). Support Vector Machines for crop/weeds identification in maize fields. *Expert Systems with Applications*, *39*(12), 11149–11155. https://doi.org/10.1016/j.eswa.2012.03.040

Iniyan, S., & Jebakumar, R. (2022). Mutual Information Feature Selection (MIFS) Based Crop Yield Prediction on Corn and Soybean Crops Using Multilayer Stacked Ensemble Regression (MSER). *Wireless Personal Communications*, *126*(3), 1935–1964. https://doi.org/10.1007/s11277-021-08712-9

Jadhav, R., & Bhaladhare, P. (2022). A Machine Learning Based Crop Recommendation System: A Survey. *JOURNAL OF ALGEBRAIC STATISTICS*, *13*(1), 426–430. https://publishoa.com

Kuanr, M., Rath, B. K., & Mohanty, S. N. (2018). Crop Recommender System for the Farmers using Mamdani Fuzzy Inference Model. In *International Journal of Engineering & Technology*. www.sciencepubco.com/index.php/IJET

Li, W., Ding, W., Sadasivam, R., Cui, X., & Chen, P. (2019). His-GAN: A histogram-based GAN model to improve data generation quality. *Neural Networks*, *119*, 31–45. https://doi.org/10.1016/j.neunet.2019.07.001

Löw, F., Michel, U., Dech, S., & Conrad, C. (2013). Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using Support Vector Machines. *ISPRS Journal of Photogrammetry and Remote Sensing*, *85*, 102–119. https://doi.org/10.1016/j.isprsjprs.2013.08.007

Madhuri, J., & Indiramma, M. (2021). Artificial Neural Networks Based Integrated Crop Recommendation System Using Soil and Climatic Parameters. *Indian Journal of Science and Technology*, *14*(19), 1587–1597. https://doi.org/10.17485/IJST/v14i19.64

Mani, D., & Edinburgh, R. (n.d.). *Crop-Yield Prediction And Crop Recommendation System*. https://ssrn.com/abstract=4111856

Mythili, K., & Rangaraj, R. (2021). Deep Learning with Particle Swarm Based Hyper Parameter Tuning Based Crop Recommendation for Better Crop Yield for Precision Agriculture. *Indian Journal of Science and Technology*, *14*(17), 1325–1337. https://doi.org/10.17485/IJST/v14i17.450

Suresh, G., Kumar, D. A. S., Lekashri, D. S., Manikandan, D. R., & Head, C.-O. (n.d.). Efficient Crop Yield Recommendation System Using Machine Learning For Digital Farming. *International Journal of Modern Agriculture*, *10*(1), 2021.

Thilakarathne, N. N., Bakar, M. S. A., Abas, P. E., & Yassin, H. (2022). A Cloud Enabled Crop Recommendation Platform for Machine Learning-Driven Precision Farming. *Sensors*, *22*(16). https://doi.org/10.3390/s22166299

Tian, L., Wang, Z., Liu, W., Cheng, Y., Alsaadi, F. E., & Liu, X. (2021). A New GAN-Based Approach to Data Augmentation and Image Segmentation for Crack Detection in Thermal Imaging Tests. *Cognitive Computation*, *13*(5), 1263–1273. https://doi.org/10.1007/s12559-021-09922-w

Ujjainia, S., Gautam, P., & Veenadhari, S. (2021). A Crop Recommendation System to Improve Crop Productivity using Ensemble Technique. *International Journal of Innovative Technology and Exploring Engineering*, *10*(4), 102–105. https://doi.org/10.35940/ijitee.D8507.0210421