1. Introduction

Crop prediction is an important component of modern agriculture as it assists farmers in selecting which crops to grow in a given climate condition, optimizing their yield and profits (Rao et al. 2022). The use of technology in agriculture has made crop prediction easier by providing more precise data climatic and weather conditions, enabling farmers to make informed decisions. Crop prediction models use various techniques such as statistical analysis, machine learning algorithms, and mathematical modeling to assist farmers in selecting a particular crop to grow in different climate conditions("S11042-023-17038-6," n.d.). Despite the advances in crop prediction, the accuracy of the models has been declining due to climate change, which has disrupted traditional weather patterns, causing changes in crucial climatic factors such as temperature, rainfall, and humidity.

Climate change refers to the changes in the Earth's atmospheric conditions over time, caused by human activities such as deforestation, industrialization, and the burning of fossil fuels(Pasquel et al. 2022). These activities have resulted in the accumulation of greenhouse gases in the atmosphere, which traps heat and causes the planet's average temperature to rise. The consequences of climate change include rising sea levels, more frequent natural disasters such as hurricanes and floods, and changes in weather patterns, leading to droughts, heatwaves, and extreme weather conditions (Taylor and Dixon 2012).

The impact of climate change on agriculture is a significant concern globally. Changes in temperature, rainfall, and humidity can significantly affect crop growth, leading to lower yields, crop failures, and food insecurity(Kalimuthu, Vaishnavi, and Kishore 2020). For example, increased temperatures can lead to shorter crop growing seasons, which can negatively affect crop yields, while droughts can lead to water shortages, further exacerbating the effects of climate change. These effects are not only limited to crops but also have consequences for livestock and fisheries, which depend on a stable environment to thrive(Xia et al. 2022).

The impacts of climate change on agriculture have led to significant challenges for farmers in accurately predicting which crops to grow, which could potentially result in substantial losses in agricultural productivity and income (Rao et al. 2022). Traditional crop prediction models based on historical weather data have become less reliable due to the unpredictable changes in weather patterns caused by climate change (Pasquel et al. 2022). As such, there is a pressing need for advanced machine learning algorithms that can accurately predict crop to grow under different climate conditions while incorporating NDVI which is used to quantify vegetation greenness and is useful in understanding vegetation density and assessing changes in plant health, providing farmers and other stakeholders in the agricultural sector with valuable insights and decision support.

The study therefore aims to develop, compare and deploy machine learning models capable of predicting crops a farmer can grow using climate conditions and satellite imagery.

2. Related works

In recent years, there has been increasing interest in utilizing machine learning and data analytics techniques to improve crop selection and yield prediction in agriculture. This section provides an overview of some of the recent works in this field.

(Reddy, Dadore, and Watekar 2019) Reddy et al. (2019) introduced a crop recommendation system to maximize crop yield in Ramtek region using machine learning. The study made use of an ensemble model with majority voting techniques using random tree, CHAID, K-Nearest Neighbor and Naïve Bayes as learner to recommend suitable crop based on factors like soil characteristics, soil types, and crop yield data. Notably, this study underscores the value of machine learning in agriculture.

(Chauhan and Chaudhary 2021) In a similar study, the paper "Crop Recommendation System using Machine Learning" by Chauhan and Chaudhary (2021) proposes an expert system that integrates IoT and ML to enable soil testing using the sensors. The sensor collected data on soil temperature, soil moisture, pH, NPK, are used in this system for monitoring temperature, humidity, soil moisture, and soil pH along with NPK nutrients of the soil respectively on a microcontroller and analyzed using machine Random Forest machine learning algorithm. Based on which, suggestions for the suitable crop to grow were made.

(Gopi and Karthikeyan 2023)Gopi and Karthikeyan (2022) presented a study on a novel multimodal machine learning based crop recommendation yield prediction (MMML-CRYP) technique. The MMML-CRYP model mainly focused on crop recommendation and crop prediction. The study employed equilibrium optimizer (EO) with kernel extreme learning machine (KELM) technique is employed for effectual recommendation of crops at the initial phase, followed by Random Forest technique which was employed for predicting crop yield accurately. The results are investigated using benchmark dataset. Experimentation outcomes highlighted the significant performance of the MMML- CRYP approach on the compared approaches with maximum accuracy of 97.91%.

(Rao et al. 2022) The paper "Crop prediction using machine learning" by Rao et al. (2022) aimed to discover the best model for crop prediction, which can help farmers to decide the type of crop to grow based on the climatic conditions and soil nutrients. K-Nearest Neighbor (KNN), Decision Tree, and Random Forest were compared using Gini and Entropy. Notably, the results revealed that Random Forest has the best performance accuracy with an accuracy of 99.32%.

While some studies have utilized machine learning algorithms to predict crop suitability, most have focused on using soil and climate data as the main predictor variable. By incorporating normalize difference vegetation index (NDVI) or satellite imagery, we aim to provide a more accurate and efficient way of predicting crop survivability.

3. Methodology and system design

This chapter describes the methodology and system design for developing and deploying machine learning models to predict crops based on climate conditions and Normalize Difference Vegetation Index. The methodology will cover the study area, research design, system architecture, operational methods, and methodological limitations.

I. Study area

The choice of Techiman Municipality in the Bono region for this study was driven by several key factors. Firstly, its recognized agricultural significance made it an ideal location for research on crop prediction and suitability. The municipality's climate variations provided a diverse dataset for model testing. Its unique agricultural challenges, such as unpredictable weather and specific crop diseases, made it a relevant study area.

Additionally, the availability of historical crop and climate data, local partnerships, and alignment with regional needs were crucial considerations. Techiman's accessibility simplified data collection and fieldwork. In summary, the selection of Techiman Municipality reflected a thoughtful evaluation of factors such as agricultural prominence, climate diversity, data availability, local collaborations, and practical considerations, making it an ideal setting for the study.

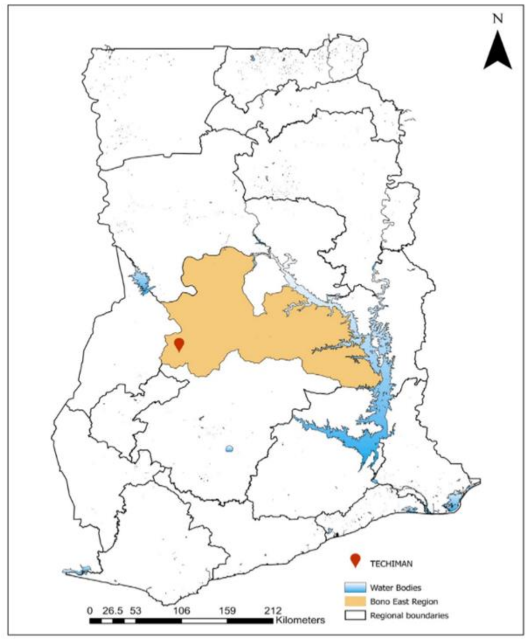


Figure 3.1 Map representation of the study area (Techiman Municipality) with red key on the map

II. Data collection

The study will use available data from the Ministry of food and Agriculture and Meteorological Department, Sunyani. The data will be obtained from the Meteorological Department, Sunyani database and will include historical climate data such as minimum temperature, maximum temperature, relative humidity at 1500, relative humidity at 0600, and rainfall, as well as crop data for selected crops, minimum NDVI, and maximum NDVI. The study will be conducted in the Techiman district located in the Bono region of Ghana.

The study will focus on selected crops grown in the Techiman district and examine how they respond to variations in climate conditions. Additionally, the study will aim to identify the most significant factors affecting crop growth, including soil conditions and climatic variables. Finally, the study findings will provide insights that can help farmers improve their farming practices and manage the risks associated with changing climate conditions.

Table 3.1 Sample of the collected dataset

Min. Temp	Max. Tem	Reltive Humidity @150	Relative Humidity@060	Rainfall	Min NDVI	Max NDVI	Crop	Code
22.5	32	42	82	7.7	0.1275	0.8474	YAM	1
22.7	35.5	38	79	44.1	0.0563	0.8513	YAM	1
23.3	34.1	50	87	79.6	0.0928	0.7488	YAM	1
23	34.2	55	88	124.3	0.157	0.8917	YAM	1
21	34	26	49	0	0.1393	0.6063	cassAVA	2
23.2	36.9	25	54	0	0.1603	0.5879	cassAVA	2
23.9	35.2	51	87	64.3	0.1559	0.6402	cassAVA	2
22.9	33.2	56	90	132.1	0.1514	0.6566	cassAVA	2
22.9	33.5	53	90	126.2	0.1306	0.5858	plantain	3
22.9	33	57	91	88	0.1384	0.4581	plantain	3
22.5	31.8	62	93	108	0.1537	0.6162	plantain	3
22	30.2	67	94	142.1	0.1568	0.6755	plantain	3
21.9	29	71	94	104.8	0.1756	0.7652	plantain	3

III. Data preprocessing

The collected data will be preprocessed to remove any inconsistencies and missing values. The preprocessing will include data cleaning, data integration, and data transformation. Data cleaning will involve removing errors and inconsistencies from the dataset. Data integration will involve combining data from different sources into a single data set. Data transformation will involve converting the data into a format that can be used by the machine learning models.

IV. Train test split

train-test split is a process in which you divide your data into a training set and a testing set. In this study, the data is divided into a 70/30 ratio, with 70% allocated to the training set and 30% to the testing set. The training set is utilized for model training, while the testing set is employed to assess the model's performance. This approach enables you to train your models using the training set and subsequently evaluate their accuracy on the unseen testing set.

V. Machine learning algorithms employed in the study

K-Nearest Neighbors (KNN): KNN is an instance-based algorithm that classifies data points by considering the majority class among their k-nearest neighbors in the feature space. This approach can be particularly valuable for a crop prediction project under varying climate conditions, as it takes into account localized climate patterns. KNN's ability to capture the similarities between data points makes it a suitable choice when considering geographical proximity and climate similarities for crop recommendations.

```
# K-Nearest Neighbor

from sklearn.neighbors import
KNeighborsClassifier

knn_classifier =
KNeighborsClassifier(n_neighbors=3)

knn_classifier.fit(X_train, y_train)
```

Decision Tree: Decision Trees are tree-like structures that utilize feature nodes to create branches representing possible outcomes. They are versatile tools for both classification and regression tasks. In the context of predicting crops under different climate conditions, Decision Trees offer interpretability and the capacity to identify nonlinear relationships between climate features and crop choices. This can help stakeholders understand which climatic factors are most influential in crop selection.

Random Forest: Random Forest is an ensemble learning method that amalgamates multiple Decision Trees to enhance model robustness and reduce overfitting. For a crop classification project dealing with potentially noisy climate data, Random Forests can be advantageous. They excel at capturing complex relationships within climate conditions, making them an appealing choice for making accurate crop predictions in diverse environmental settings.

```
# Random Forest
from sklearn.ensemble import
RandomForestClassifier
RF = RandomForestClassifier(n_estimators=20,
random_state=0)
RF.fit(Xtrain,Ytrain)
```

Logistic Regression: Logistic Regression is a straightforward linear model suitable for binary and multi-class classification. It models the likelihood of a sample belonging to a specific class. In is a feasiblee there is a discernible linear separation in climate data influencing crop choices, Logistic Regression offers simplicity and interpretability, making it a feasible option for crop prediction tasks.

```
# Logistic regression
from sklearn.linear_model import
LogisticRegression
 LogReg =LogisticRegression(random_state=2)
LogReg.fit(Xtrain, Ytrain)
```

Support Vector Machine (SVM): SVM is a potent classification algorithm that seeks to identify a hyperplane that optimally separates different classes while maximizing the margin between them. In scenarios involving high-dimensional climate data and the need to handle nonlinear relationships, SVM can be highly effective. It employs kernel functions to adapt to intricate climate-crop relationships, making it well-suited for complex prediction tasks.

```
# Support Vector Machine(SVM)
from sklearn.svm import SVC
SVM = SVC(gamma='auto')
SVM.fit(Xtrain, Ytrain)
```

AdaBoost: AdaBoost is an ensemble learning technique that combines multiple weak learners (often Decision Trees) to create a robust classifier by assigning higher weights to misclassified samples. When dealing with complex climate-crop relationships, AdaBoost can adapt by focusing on the challenging-to-classify data points, thereby enhancing prediction accuracy.

```
# AdaBoost

from sklearn.ensemble import AdaBoostClassifier

from sklearn.tree import DecisionTreeClassifier

from sklearn.datasets import load_iris

from sklearn.model_selection import train_test_split

from sklearn.metrics import accuracy_score

weak_classifier = DecisionTreeClassifier(max_depth=1)

adaboost_clf = AdaBoostClassifier(base_estimator=weak_classifier, n_estimators=50, random_state=42)

adaboost_clf.fit(X_train, y_train)
```

Soft Voting:

Soft voting, also known as weighted voting, is a technique used in ensemble learning, particularly with classification algorithms. In soft voting, multiple models (classifiers or regressors) make predictions on a given input, and instead of simply counting the majority vote (as in hard voting), the models' predictions are combined by taking into account their confidence scores or probabilities.

```
# Soft voting

from sklearn.ensemble import VotingClassifier

from sklearn.tree import DecisionTreeClassifier

from sklearn.neighbors import KNeighborsClassifier

from sklearn.svm import SVC

clf1 = DecisionTreeClassifier(random_state=42)

clf2 = KNeighborsClassifier(n_neighbors=5)

clf3 = SVC(probability=True, kernel='linear')

voting_clf = VotingClassifier(estimators=[('decision_tree', clf1), ('knn', clf2), ('svm', clf3], voting='soft')

# Train the ensemble model

voting_clf.fit(X_train, y_train)
```

Hard Voting:

Hard voting is a simple ensemble learning technique in machine learning. In hard voting, multiple machine learning models, such as classifiers, are trained on the same dataset. When it's time to make a prediction, each model "votes" for a class, and the class that receives the majority of votes is selected as the final prediction. Essentially, it's like taking a popular vote among the models to decide the prediction. Hard voting works best when the individual models are diverse and perform reasonably well on their own, as it leverages the wisdom of the crowd to make more accurate predictions.

```
# Hard voting

from sklearn.ensemble import VotingClassifier

from sklearn.tree import DecisionTreeClassifier

from sklearn.neighbors import KNeighborsClassifier

from sklearn.svm import SVC

clf1 = DecisionTreeClassifier(random_state=42)

clf2 = KNeighborsClassifier(n_neighbors=5)

clf3 = SVC(probability=True, kernel='linear')

voting_clf = VotingClassifier(estimators=[('decision_tree', clf1), ('knn', clf2), ('svm', clf3], voting='hard')

voting_clf.fit(X_train, y_train)
```

Stacking Classifier

Stacking classifier, or stacking ensemble, is a machine learning technique that combines multiple base models (classifiers or regressors) with a meta-model to improve predictive performance. Stacking is a form of ensemble learning where the predictions of the base models are used as input features for a higher-level model, which makes the final predictions. Stacking helps to leverage the strengths of various base models, compensating for their weaknesses, and potentially achieving better predictive accuracy.

```
# Stacking classifier voting
from sklearn.ensemble import StackingClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.neighbors import
KNeighborsClassifier
from sklearn.svm import SVC
base_classifiers = [
    ('knn', KNeighborsClassifier(n_neighbors=5)),
    ('svm', SVC(kernel='linear', probability=True))
]
stacking_clf =
StackingClassifier(estimators=base_classifiers,
final_estimator=KNeighborsClassifier(n_neighb
ors=3))
stacking_clf.fit(X_train, y_train)
```

In sum, these models collectively contribute the growing role of machine learning and technology in improving various fields including agriculture, from crop selection, and underscore their potential to drive sustainable practices and enhance food security.

Table 3.1 Parameter summary of machine learning model implementation

Machine Learning Models	Parameters
Knearest Neigbor	n_neighbors=2, metric= "minkowski"
Decision Tree	criterion = "gini", min_samples_leaf=2, random_state = 0
Random Forest	Criterion="gini",in_samples_leaf=2,n_estimators=10, random_state =0
Support Vector Machine	kernel="linear", random_state = 0
Logistic Regression	()
Adaboost	base_estimator=RFC
Voting 1	Estimators="DTC",dtc, "rf",RFC, "knn",knn,voting='soft', weights=None, n_jobs=None, flatten_transform=True,
Voting 2	Estimators="DTC",dtc,"rf",RFC,"knn",knn, voting='hard', weights=None, n_jobs=None, flatten_transform=True,
Stacking	Estimators=" DTC",dtc, "rf",RFC, "knn",knn,

The random state value is set to zero, as shown in the table (0). In essence, the random state is an instance of NumPy that is an integer. Which random number generator to use is specified by RandomState, None, or (default). In a sense, the zero value serves as a model's default value. With the same random state value set to zero, the decision tree also used gini as a criterion. A number of 10 estimators were used by the random forest classifier. The neighbor selection for KNN is five, and the distance measure used is the minkowski's distance. Two variations of SVM was used were the radial basis function and linear technique.

VI. Evaluation metrics

Evaluation metrics are essential tools in this study as they provide a systematic and quantitative means to assess the performance and accuracy of machine learning models in predicting crop choices under varying climate conditions. These metrics enable researchers and stakeholders to gauge how well the models align with real-world outcomes and make informed decisions. In the context of crop prediction, the selection of the appropriate evaluation metrics ensures that the recommendations provided by the models are not only reliable but also tailored to the specific needs of farmers and agricultural planners. By measuring factors such as accuracy, precision, recall, F1-score, and area under the receiver operating characteristic curve (AUC-ROC), we can quantitatively determine the models' ability to correctly classify crops based on climate data, helping to optimize agricultural practices, enhance sustainability, and adapt to changing environmental factors.

VII. Confusion metrics

According to (Li, Sun, and Li 2023) a confusion matrix bears details about actual and predicted classifications been used by a classification system. Confusion matrices consist of two dimensions actual and predicted value (Shultz et al., 2011). Simply an M x M matrix. Where M is the number of classes being predicted.

Table 3.3.2 confusion matrix presentation

		Predicted outcome	
		Negative	Positive
Actual outcome	Negative	TP	FP
	Positive	FN	TN

Accuracy: (AC) This statistic determines the percentage of correctly predicted outcomes across all cases. The formula below is used to estimate.

Equation 1

$$(\mathrm{AC}) = \frac{\mathrm{TP+TN}}{\mathrm{TP+FP+FN+TN}}$$

Recall or the True Positive Rate (TP): this refers the proportion of positive cases that were correctly identified, as calculated using the formula below.

Equation 2

$$(\mathrm{TP}) = \frac{\mathrm{FN}}{\mathrm{FN+TN}}$$

The False Positive Rate (FP): This reflects the proportion of incorrectly labeled negative cases. Using the output of the formula.

Equation 3

$$(\mathrm{FP}) = \frac{\mathrm{FP}}{\mathrm{TP+FP}}$$

True Negative Rate (TN) this is the percentage of incorrectly classified negative cases as determined by the formula below.

Equation 4

$$(\mathrm{TN}) = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$$

The percentage of negative instances that were mistakenly labeled as negative is known as the false negative rate (FN). According to this formula

Equation 5

$$(\mathrm{FN}) = \frac{\mathrm{FN}}{\mathrm{FN+TN}}$$

Precision (P) this is the percentage of correctly predicted positive situations that were determined using the formula.

Equation 6

$$(P) = \frac{TN}{FP+TN}$$

Classification report

The Sci-learn library offers a selection of practical reporting tools for handling classification issues so that one can get a solid understanding of the model's accuracy considering a variety of criteria. The classification_report () function is provided on the Jupyter notebook (anaconda3) platform uses a set of algorithms to return the following specific metrics in the evaluation of machine learning models. They include Precision, Recall, F1 score, and Support.

In essence, the listed evaluation metrics are considered in the evaluation of the used machine learning model in this research. As hinted earlier, this methodology is to give a better understanding of the behavior of these machine learning algorithms for the climatic weather conditions and crop(s) dataset understudy.

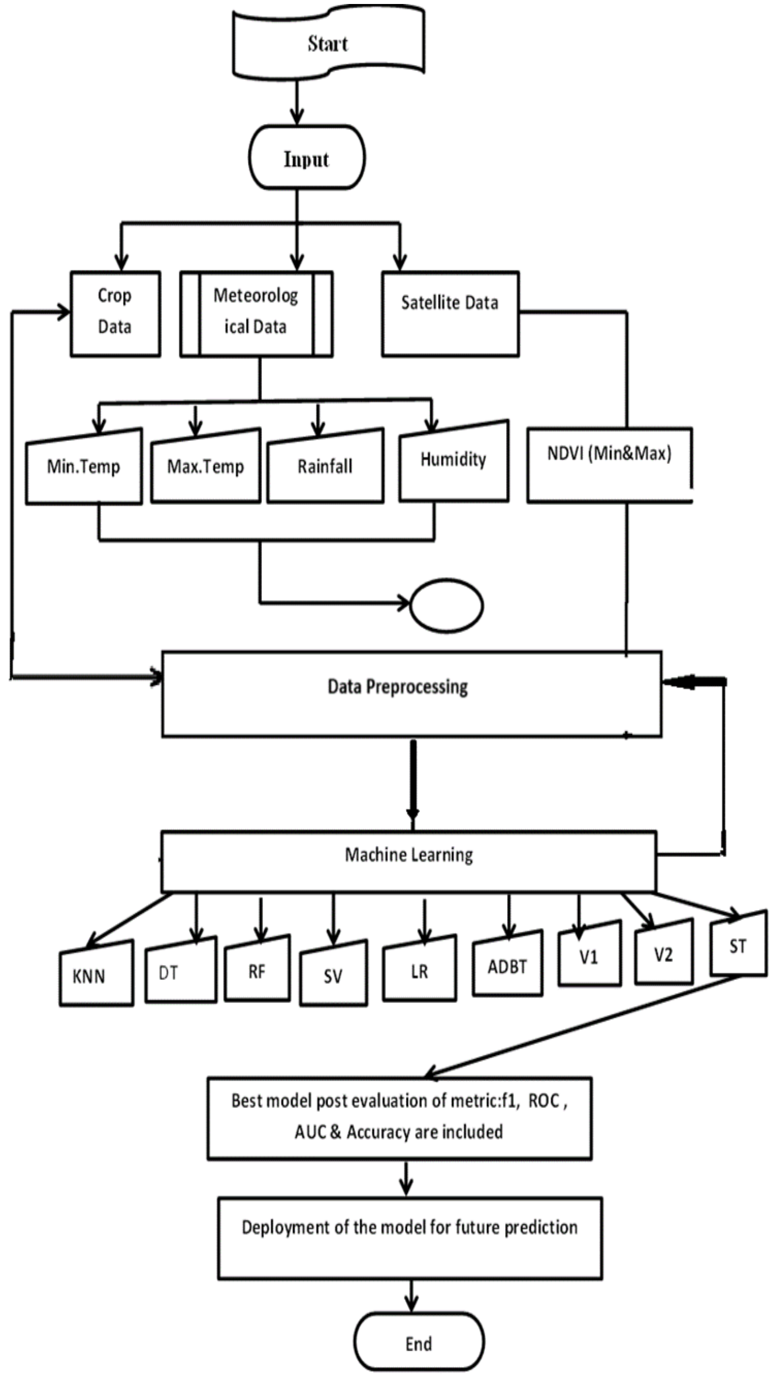
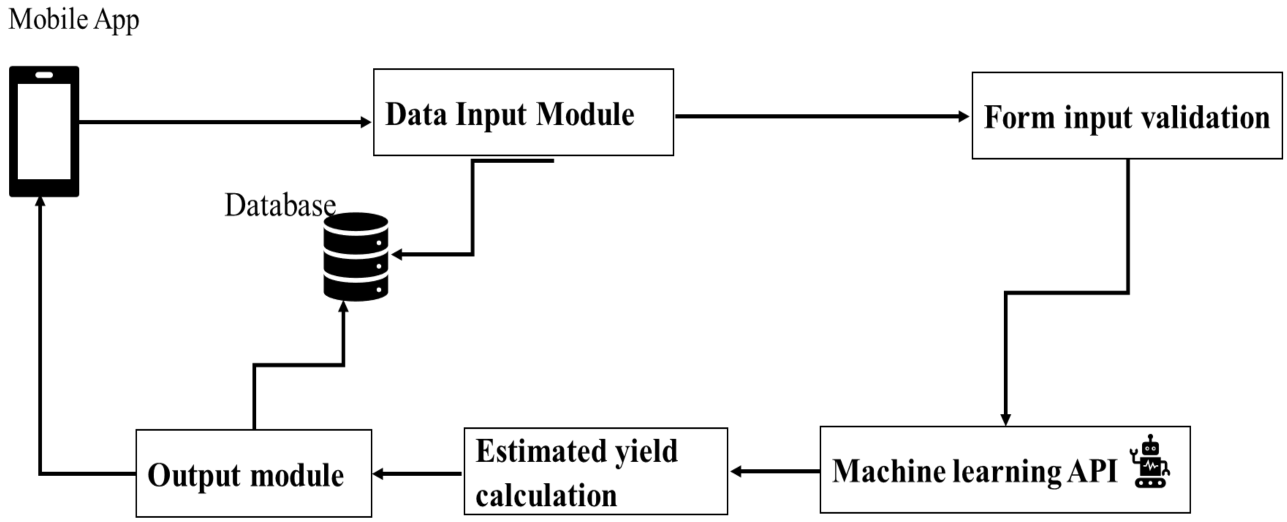


Figure 3.2 Framework of Machine learning process for the study

VIII. Overview of the proposed system



The overview of the proposed system highlights how the important component within the mobile application communicates with machine learning API to predict crops that a farmer can grow and also provide the estimated yield of the predicted crop.

Data input module: The data input module will accept climate and NDVI data from the user, preprocess the data, and transform it into a format that can be used by the machine learning model. This module will be developed using java and will utilize the tablesaw package for data processing.

Form input validation: The form input validation function validates the data entered by the user before the input module preprocesses the data.

Machine learning API: The machine learning model module API is python flask server hosted on a python anywhere cloud server which contains the machine learning model with the highest performance accuracy saved as pickle file.

Estimated yield calculation: The estimated yield calculation is a function that calculates the estimated yield of the predicted crop by taking into consideration, the land size, estimated number of fruits on each plant and the spacing between each plant before sending them to the output module for display.

Output Module: The output module will generate the predicted optimal crop for the given environment based on the input data. This module will be developed using Python and will output the results in a user-friendly format on the mobile app UI.

Database: A database will be used to store the weather data entered by the app users, as well as the output generated by the machine learning model. The database will be implemented using SQLite, a powerful open-source relational database management system.

4. Results and discussion

I. Overview of the exploratory data analysis

A visual representation of our sample dataset through screenshots, a correlation matrix illustrating the pairwise relationships among features, and a concise summary of the parameters used in our machine learning model implementations. These elements collectively offer a comprehensive glimpse into our research methodology and data analysis.

Screenshot of sample dataset, Correlation matrix-showing pairwise correlation among features, Parameter summary of machine learning model implementation

	MIN_TEMP	MAX_TEMP	RH150	RH60	Rainfall	Area cropped(Ha)	Yield (Mt/Ha)	MIN_NDVI	MAX_NDVI	CROP
0	22.5	32.0	42.0	82	7.7	6820	17.65	0.1275	0.8474	1
1	22.7	35.5	38.0	79	44.1	5047	18.90	0.0563	0.8513	1
2	23.3	34.1	50.0	87	79.6	7852	2.40	0.0928	0.7488	1
3	23	34.2	55.0	88	124.3	1704	9.80	0.1570	0.8917	1
4	22.8	32.6	61.0	91	154.7	276	9.96	0.0749	0.8650	1

Figure 4. 1 Screenshot of sample dataset

Figure 4. 2 representation of correlation matrix using heatmap

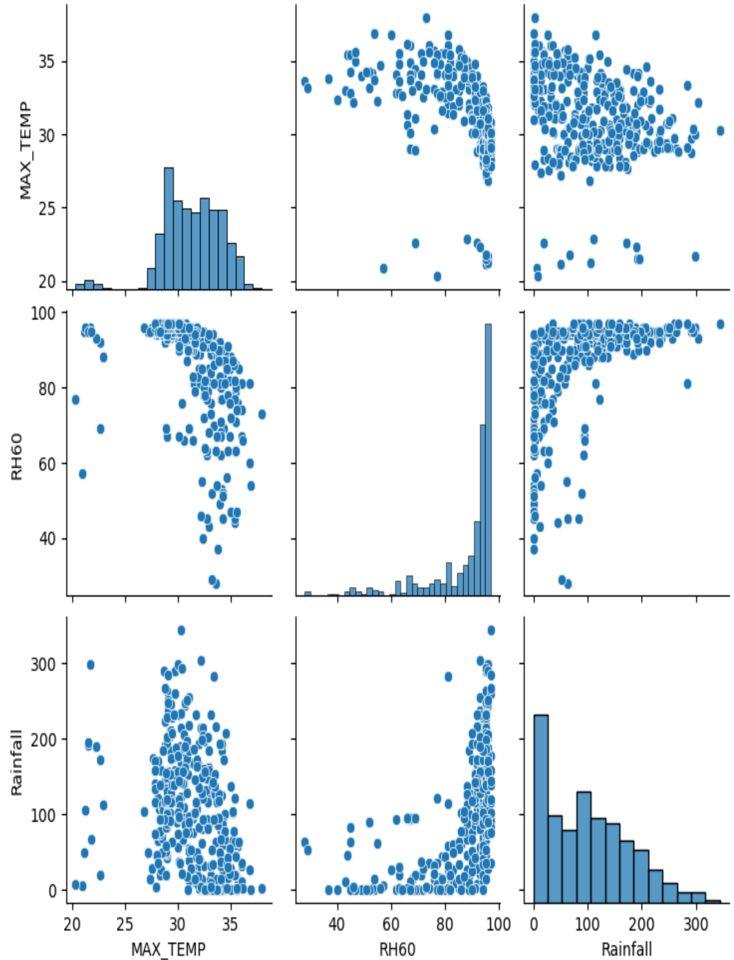


Figure 4. 3 Data Distribution Plot for Some of the Features

II. Results discussion

Before considering evaluation metrics, the machine learning model's training and testing accuracy must first be generated. A summary of the training and testing accuracies attained after implementation is provided in Table 4.2 (training and testing accuracies of the various machine learning models.)

A wealth of tools and resources are available in the Scikit library to test run interesting algorithms.

Table 4.1 Various machine learning models with their training and testing accuracy

Machine learning model	Testing accuracy	Training accuracy
K-nearest Neighbor	0.68	0.73
Decision Tree	0.83	0.98
Random Forest	0.85	0.95
Support Vector Machine	0.55	0.55
Logistic regression	0.61	0.63
AdaBoost	0.57	0.59
Voting 1(hard)	0.85	0.91
Voting 2 (soft)	0.86	0.93
Stacking	0.85	0.92

The testing and training accuracies were computed through the use of the confusion matrix using Jupiter notebook (anaconda).

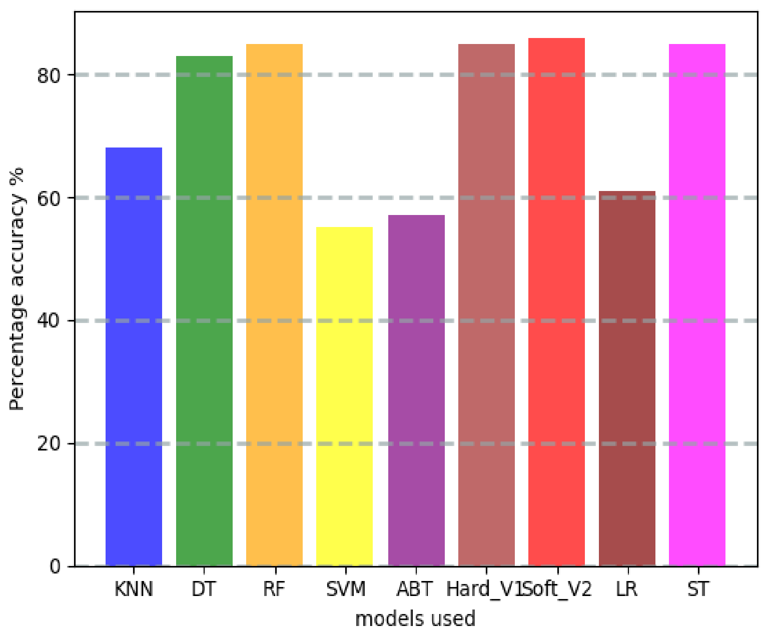


Figure 4.4 Accuracy Comparison among All Classifiers for Specific Crop Possibility

Table 4.2 Statistical measures for the validation of machine learning algorithms.

Algorithms.	**MSE**	**RMSE**	**R^2**
K-nearest neighbor	0.259	0.509	-0.045
Decision Tree	0.152	0.508	0.385
Random Forest	0.107	0.326	0.569
Support Vector Machine	0.431	0.657	-0.742
Stack Ensemble	0.152	0.509	0.385
Harding Voting	0.446	0.668	-0.804
Soft Voting	0.101	0.319	0.590
AdaBoost	0.244	0.494	0.016
Logistic Regression	0.437	0.660	-0.762

Figure 4. 5 Graph of F1 Score and Precision among all the classifiers

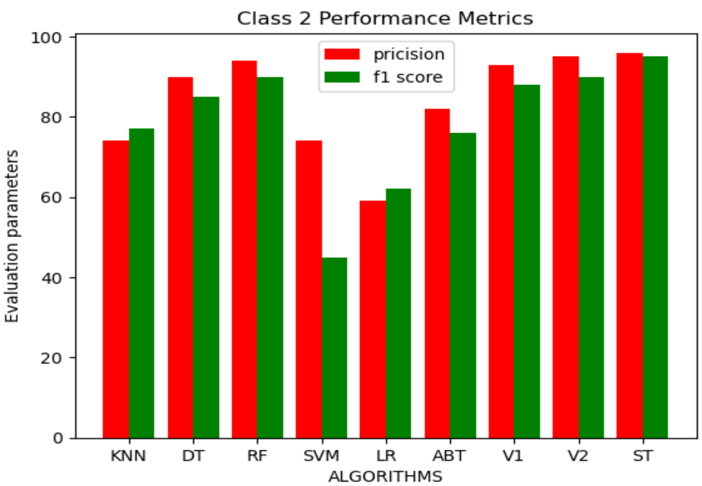


Figure 4. 6 Graph of F1 Score and Precision among all the classifiers

III. Discussion

The study assessed five different machine learning models and four ensemble algorithms for predictive capabilities including K-nearest neighbors (KNN), Support Vector Machine, Decision Tree, Logistic Regression, Random Forest, Soft Voting, AdaBoost, Stacking Classifier, and Hard Voting. Among the nine models, Decision Tree, Random Forest, Hard Voting, Soft Voting and Stacking classifier obtained high performance accuracy of 82.90%, 84.71%, 85.10%, 85.77%, and 85.12% respectively with the soft voting model having the highest performance accuracy of 85.77% as shown in **figure 4.4**. The soft voting model was then saved as pickle file and imported into the flask server for hosting where communication was established between the mobile application and the flask server for testing. During the testing phase, the mobile application was able to predict based on the dataset entered in the “predict crop form”. The data used for the testing were collected from the study’s dataset and from reputable sources such as NASA POWER.

This configuration enabled real-time predictions to be generated based on user-input climate conditions. Subsequently, the server's deployment transitioned to a more robust infrastructure on Heroku. By doing so, the application's availability and reliability were significantly enhanced, allowing a broader user base to access the crop prediction service seamlessly. This transition to Heroku hosting exemplifies the commitment to delivering dependable and accessible services to users seeking informed decisions regarding crop cultivation in varying climatic contexts. Below are some of the images from the mobile application.

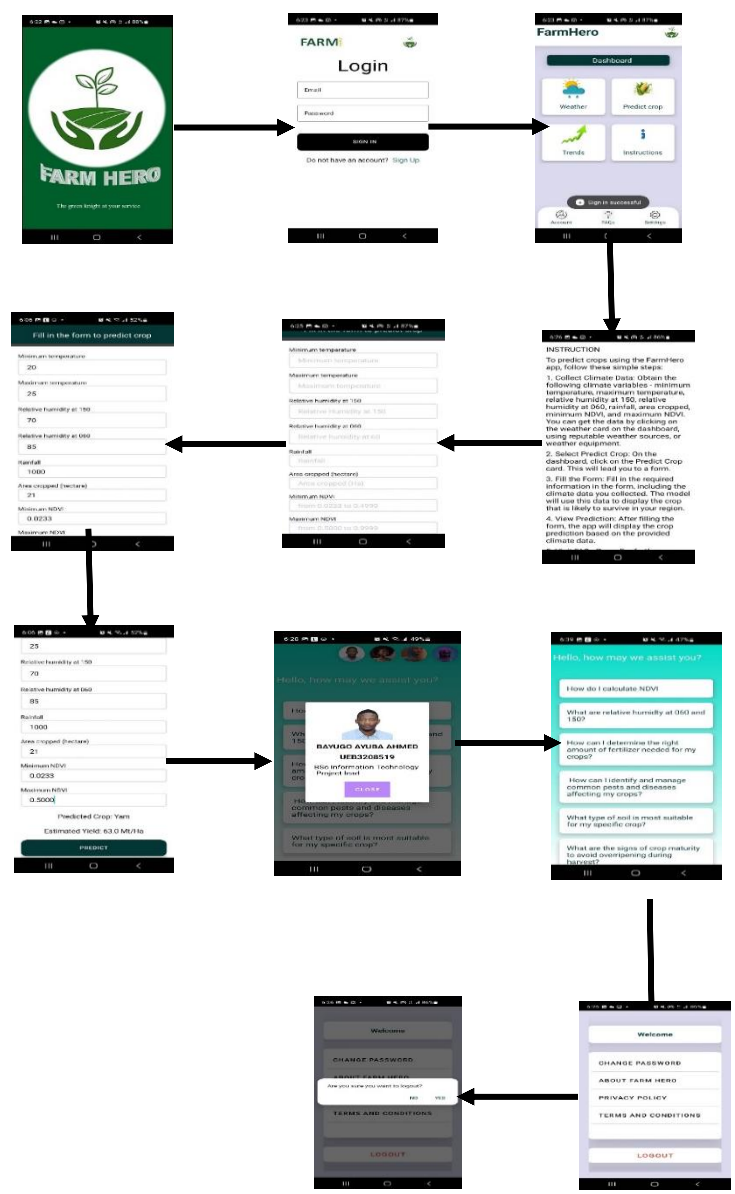


Figure 4.7 Images of the mobile application features

IV. Conclusion

The study demonstrates the effectiveness of machine learning models in predicting crops a farmer can grow based on climate conditions and normalize difference vegetation index. Hyperparameter tuning enhances accuracy, and models like Decision Tree, Random Forest, and Voting Classifiers perform notably well. The successful integration of these models into a mobile app marks a promising step toward informed agricultural decisions, though regional variations should be acknowledged. This research advances precision agriculture through machine learning's potential to optimize crop yield.

V. Future work

Base on the study conducted on the 9 machine learning models, it can be recommended that;

- Incorporating real-time weather data: Future research can explore the integration of up-to-date weather information to improve the accuracy of crop suitability assessments. This would enable the models to adapt dynamically to changing climate conditions.
- Future research endeavors can consider the integration of Internet of Things (IoT) technology to further enhance predictive capabilities.
- Future studies may also benefit from exploring the impact of fertilizer application on crop suitability and performance.
- Expanding the dataset's scope by incorporating information about soil compositions is recommended. This addition can help prevent land from going to waste in the era of precision agriculture, ensuring more informed land-use decisions.

References

Chauhan, Gaurav, and Alka Chaudhary. 2021. “Crop Recommendation System Using Machine Learning Algorithms.” *Proceedings of the 2021 10th International Conference on System Modeling and Advancement in Research Trends, SMART 2021* 3307: 109–12. https://doi.org/10.1109/SMART52563.2021.9676210.

Gopi, P. S.S., and M. Karthikeyan. 2023. “Multimodal Machine Learning Based Crop Recommendation and Yield Prediction Model.” *Intelligent Automation and Soft Computing* 36 (1). https://doi.org/10.32604/iasc.2023.029756.

Kalimuthu, M., P. Vaishnavi, and M. Kishore. 2020. “Crop Prediction Using Machine Learning.” *Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020*, no. Icssit: 926–32. https://doi.org/10.1109/ICSSIT48917.2020.9214190.

Li, Jingzheng, Hailong Sun, and Jiyi Li. 2023. “Beyond Confusion Matrix: Learning from Multiple Annotators with Awareness of Instance Features.” *Machine Learning* 112 (3): 1053–75. https://doi.org/10.1007/s10994-022-06211-x.

Pasquel, Daniel, Sébastien Roux, Jonathan Richetti, Davide Cammarano, Bruno Tisseyre, and

James A. Taylor. 2022. “A Review of Methods to Evaluate Crop Model Performance at Multiple and Changing Spatial Scales.” *Precision Agriculture* 23 (4): 1489–1513. https://doi.org/10.1007/s11119-022-09885-4.

Rao, Madhuri Shripathi, Arushi Singh, N. V.Subba Reddy, and Dinesh U. Acharya. 2022. “Crop Prediction Using Machine Learning.” *Journal of Physics: Conference Series* 2161 (1): 23–26. https://doi.org/10.1088/1742-6596/2161/1/012033.

Reddy, D. Anantha, Bhagyashri Dadore, and Aarti Watekar. 2019. “Crop Recommendation System to Maximize Crop Yield in Ramtek Region Using Machine Learning.” *International Journal of Scientific Research in Science and Technology* 6 (1): 485–89. https://doi.org/10.32628/ijsrst196172.

“S11042-023-17038-6.” n.d.

Taylor, Publisher, and Geoffrey Richard Dixon. 2012. “Climate Change – Impact on Crop Growth and Food Production , and Plant Pathogens Symposium Contribution / Contribution à Un Symposium Climate Change – Impact on Crop Growth and Food Production , and Plant Pathogens.” *Canadian Journal of Plant Pathology* 34 (3): 362–79.

Xia, Lili, Alan Robock, Kim Scherrer, Cheryl S. Harrison, Benjamin Leon Bodirsky, Isabelle Weindl, Jonas Jägermeyr, Charles G. Bardeen, Owen B. Toon, and Ryan Heneghan. 2022. “Global Food Insecurity and Famine from Reduced Crop, Marine Fishery and Livestock Production Due to Climate Disruption from Nuclear War Soot Injection.” *Nature Food* 3 (8): 586–96. https://doi.org/10.1038/s43016-022-00573-0.