A FIELD PROJECT REPORT

on

“Resilient Farming with Climate-Based Crop Guidance”

Submitted

by

221FA04128

Yaswanth Krishna Kumar Pothuri

221FA04096

Abhirama Raju Nadimpalli

221FA04421

Prasanth Tuta

221FA04156

Venkata Naga Sai Kiran Kothuru

221FA04451

Prince kumar

Under the guidance of

Bhargavi Maridu

Assistant Professor Department of CSE, VFSTR

VIGNAN'S FOUNDATION FOR SCIENCE, TECHNOLOGY & RESEARCH

(Deemed to be University) - Estd. u/s 3 of UGC Act 1956

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

VIGNAN'S FOUNDATION FOR SCIENCE, TECHNOLOGY AND RESEARCH

Vadlamudi, Guntur.

ANDHRA PRADESH, INDIA, PIN-522213.

CERTIFICATE

This is to certify that the Field Project entitled **"Resilient Farming with Climate-Based Crop Guidance"** that is being submitted by 221FA04128 (Yaswanth Krishna Kumar Pothuri), 221FA04096(Abhirama Raju Nadimpalli), 221FA04421 (Prasanth Tuta), 221FA04156(Venkata Naga Sai Kiran Kothuru) and 221FA04451(Prince kumar) for partial fulfilment of Field Project is a bonafide work carried out under the supervision of Ms. Bhargavi Maridu , M.Tech., Assistant Professor, Department of CSE.

M. Bhargavi	Dr. S.V. Phani Kumar	Dr.K.V. Krishna Kishore
Assistant Professor, CSE	HOD,CSE	Dean, SoCI

DECLARATION

We hereby declare that the Field Project entitled **"Resilient Farming with Climate-Based Crop Guidance"** that is being submitted by 221FA04128 (Yaswanth Krishna Kumar Pothuri), 221FA04096(Abhirama Raju Nadimpalli),221FA04421(Prasanth Tuta),221FA04156(Venkata Naga Sai Kiran Kothuru) and 221FA04451(Prince kumar) in partial fulfilment of Field Project course work. This is our original work, and this project has not formed the basis for the award of any degree. We have worked under the supervision of Ms. Bhargavi Maridu, M.Tech., Assistant Professor, Department of CSE.

By

221FA04128 (Yaswanth Krishna Kumar Pothuri),

221FA04096(Abhirama Raju Nadimpalli),

221FA04421(Prasanth Tuta),

221FA04156(Venkata Naga Sai Kiran Kothuru),

221FA04451(Prince kumar)

Date:

ABSTRACT

Agriculture is the main source of income in rural India and has been positively contributing to the nation's GDP. However, the crop produced per hectare in this country is less than the global average, causing misery to farmers, who even take their own lives the marginal farmers are no exception.This research seeks to tackle these challenges by developing a climatebased crop recommendation system powered by machine learning.The system uses critical field inputs, including temperature, soil moisture, nutrient levels (NPK), pH, and rainfall, the system indicates potential crops to help farmers make databased decisions about using the right crop choice Of the three machine learning algorithms taken-Decision Tree, Random Forest, and Logistic Regression, the highest accuracy was found in Navie Bayes, which reached a peak of 99.54%.Besides this, farm practice also has changed with technological advancement. Mechanization and precision agriculture ensure that crop production yields quality and increases yield. A further development of the model provides a machine learning model that relies upon climatic variability and recorded weather to generate recommendations specific to the location for crops. Intermingling weather patterns with crop performance leads to better yields and sustainable agriculture practices. Future iterations of this could include other variables like soil conditions and market Index Terms—Data pre-processing, Min-Max scaler,Grid search ,Machine Learning, Logistic Regression, Random Forest, XG-Boost Algorithm Algorithm, Accuracy.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1. INTRODUCTION

1.1 Motivation

Agriculture remains the backbone of the rural economy in India and significantly contributes to the nation's GDP. However, the crop yield per hectare in India lags behind the global average, which has led to challenges for farmers, particularly small-scale and marginal farmers. This situation has, at times, led to dire outcomes, including farmers taking their own lives due to financial distress. With advancements in technology, there is an opportunity to address these issues by providing farmers with data-driven insights to make more informed decisions about crop selection. The motivation behind this project is to develop a system that can offer resilient farming solutions by utilizing climate-based crop guidance, thereby improving agricultural productivity and supporting farmers' livelihoods.

1.2 Problem Definition/Research Gaps

Traditional farming practices often rely on experience and general advice, which may not account for specific local environmental conditions and changes. This leads to inefficient use of resources and suboptimal crop yields. While some tools exist for crop recommendation, they may not comprehensively consider critical factors such as soil nutrients, climate variability, and weather patterns. This project aims to bridge this gap by developing a climate-based crop recommendation system that integrates environmental and soil parameters to suggest the most suitable crops. By applying machine learning techniques, the system will predict potential crops, helping farmers to optimize their practices and improve crop yield sustainably.

1.3 Limitations

The following limitations are taken into consideration for designing a precise and reliable crop recommendation model:

- **Dataset Quality**: The accuracy of the model depends on the quality of the input data. Incomplete or biased data can lower prediction performance.
- **Computational Power**: Training complex models may require significant computational resources, which could be a challenge for extensive datasets.
- **Overfitting**: Models like decision trees can be prone to overfitting without proper regularization, which affects performance on new data.

- **Interpretability**: While simpler models are easy to understand, more complex models may act as "black boxes," making it difficult to interpret how predictions are made.

1.4 Design Standards

The proposed predictive model will utilize machine learning techniques, adhering to industry standards for accuracy, efficiency, and ethical considerations. The model will consider various factors such as soil characteristics, weather data, and previous crop performances to make precise recommendations, helping farmers make data-informed decisions.

1.5 Major Contributions / Objectives

The primary objective of this project is to develop a robust and efficient climate-based crop recommendation model that:

- Utilizes **machine learning algorithms** to integrate environmental and soil data for accurate crop predictions.
- Applies **Principal Component Analysis (PCA)** for dimensionality reduction to simplify data without losing essential information.
- Employs **classification and regression techniques** such as Random Forest, Gradient Boosting, and Naive Bayes for effective model training.
- **Compares performance** of traditional and advanced machine learning models to identify the most effective approach.
- Provides farmers with **real-time crop recommendations** that can be used to make informed agricultural decisions, thereby promoting sustainable farming practices.

CHAPTER 2

LITERATURE SURVEY

2. LITERATURE SURVEY

2.1 Literature review

A literature survey is a systematic examination of existing research on a particular topic. It serves as the foundation for any scholarly investigation, offering insights into current knowledge, identifying research gaps, and providing context for new studies. By synthesizing and summarizing relevant literature, researchers can formulate precise research questions, build upon existing work, and avoid duplication. In essence, a literature survey is an essential tool for ensuring the validity and relevance of new research within the broader academic landscape.

Table 1: Literature survey

No	**Author(s)**	**Model/Approach**	**Accuracy/Results**	**Limitation**
1	Pachade, R. S., & Sharma, A.	Decision Tree, Random Forest, and Logistic Regression	Random Forest Regression achieved 99.32% accuracy	Improvements aim to incorporate more crop data and additional soil nutrient parameters for enhanced precision
2	Shams, M.Y., Gamel, S.A., & Talaat, F.M.	Decision tree, Bagging ,AdaBoost ,Gradient boosting ,Random forest	Random forest 0.9188 accuracy	While the current model provides a single crop label recommendation,

3	Mahale, Y., Khan, N., Kulkarni, K., et al.	Decision tree, Bagging, AdaBoost, Gradient boosting, Random Forest	Random forest 0.9188 accuracy	While the current model provides a single crop label recommendation, future enhancements could explore presenting additional information.
4	Gosai, D., Raval, C., Nayak, R., Jayswal, H.,& Patel, A.	XGBoost and Improved Learning Model	Reduced risk of overfitting with large dataset	Overfitting with smaller datasets

Pachade, R. S. [1] further claims that these are now finding increased usage in agricultural yield predictions of crops and pest detection. These include methods like decision trees and neural networks that scan large databases with much higher accuracy than traditional methods. Growth of crops is dependent upon the weather conditions; therefore, weather-based models are developed. While earlier systems were general crop recommendation systems, the current ML-based system considers particular weather, soil, and historical information to provide a relevant recommendation, being dynamic and assisting farmers in making an informed decision.

Shams, M.Y., Gamel, S.A. & Talaat [2] Crop recommendation systems can include recent advancements in using machine learning to help farmers with decisionmaking. For example, Shams et al. (2024) proposed the XAI-CROP, where the algorithm is also transparent, primarily based on the use of the principles of eXplainable Artificial Intelligence (XAI). Trained on Indian crop data, XAI-CROP utilizes an integrated decision tree model that employs Local Interpretable Model-agnostic Explanations (LIME) to give understandable recommendations. It achieved a higher accuracy with a low MSE of 0.9412 and an R2 value of 0.94152, thus enhancing the transparency in the way AI-driven agricultural decisions are made. This research underscores the role of interpretation in machine learning models.

Mahale et al. [3] have also designed a crop recommendation and forecasting system for Maharashtra state by using machine learning techniques, such as the combination of LSTM models with a novel expectation-maximization approach. The system increases the accuracy of crop predictions with multifactorial approaches in agriculture, including climatic and soil settings. This research highlights the requirements for advanced machine learning methodology in crop yield advancement and effective farmers' decision-making. To learn more, you can read the full article.

In [11], authors proposed a smart way to manage crops and harvest them. A variety of machine learning models like KNN, Logistic Regression, Bagging, Naïve-Bayes, SVM, AdaBoost, Decision Tree, RF, Gradient-Boosting, XGB, IBGM were used for crop recommendation where Random Forest gave better accuracy of 97.18

An ensemble model named KKR (Korringa Kohn Rostoker) for proper crop cultivation in Bangladesh [12] uses distance-based KNN' &' second-order ensemble strategy which finds the combined predictors from KNN, RR, and RF to form an ensemble regressor predicting better. The proposed KKR performed better to investigate three major rice variations with potatoes and help authorities planning the food supply in the future. However, only the major crops are predicted. There were some factors that were omitted from the analysis, such as soil properties, production costs, and market prices. Comparisons of finely tuned machine learning models explain the variability of wheat yields on Indian farmlands in the northwest Indo-Gangetic Plains has been made in [13]. The random forest model showed good fitting, accuracy and precision measures, which accounted for 25% of wheat yield variability. Future research is required to understand the relation between crop yield for prioritizing and ensuring sustainable farming. Authors in [14] conducted an in-depth literature analysis for crop prediction with ML. Around 567 of relevant studies were retrieved, of which 50 were selected for further analysis. The authors conclude that temperature rainfall, and soil variety are the commonly used features, and the most common algorithm used is CNN. In paper [15], the authors experimented with 5 different ML algorithms viz Decision Tree, Naive Bayes, SVM, Random Forest and Logistic Regression for West Bengali Agriculture focusing on 13 major crops for different districts in West Bengal. Rainfall, Temperature, Humidity, Sun Hours were considered, and it was observed that RF and SVM helped to achieve superior results. Although the project is tested on 2 district datasets only, work needs to be done on updating the databases for error-free predictions. In [16] the authors proposed a prediction model for rice cultivation with Support Vector Machine in China. PCA and fivefold cross-validation were used for dimensional reduction with three types of rice plantings and executed successfully. In addition to being convenient for data acquisition, the proposed open-crop model integrates multiscale factors well, is simple to parametrize and applies to a wide range of regions. The model can be further expanded to include variables like the cost of grain, fertilizer, seed, labour, location, traffic, support from the government, the true state of agriculture, and innovations in science and culture. A deep learning model using CNN was studied in [17]. They identified infections in plants and studied them to prevent loss of crop prediction and help farmers grow healthy plants. This method increased the efficiency compared to traditional methods as it uses image processing. This technique can be modified by considering climatic factors and considering more diseases in plants. Although the system is currently only implemented in Karnataka, it can be expanded to cover the entire country. In [18] Data Analytics techniques such as Linear Regression with neural networks has been used for prediction of prices for crops. Factors such as Area harvested, Area planted, etc. were considered and the authors concluded that XGBoost was the best technique for price prediction of various crops while in [19] authors predicted crop cultivation for farmers using machine learning techniques which suggests best conditions to plant, harvest, and water crops. They used pair plots, joint plots, heat maps, and Barh for decision tree prediction. Moreover, the models can be modified for fertilizer recommendation and app integration, so that farmers can have easy access to them. For fertilizer utilization and crop prediction, a machine learning regression algorithm based on naive Bayes classifiers was developed in [20]. A model was developed based on 4 crops and 7 parameters, including soil nutrients, in the Mysore district. For crops such as wheat, ragi, and paddy, the algorithm has a good prediction rate. Furthermore, it can also be modified through the use of farmer-friendly applications. The study in [21] delves into the evolving nature of Indian Summer Monsoon (ISM) variability in Maharashtra, Western India, and its impact on agriculture and food security. Unlike previous coarse-scale analyses, this research focuses on a finer district-level examination, revealing significant Spatio-temporal heterogeneity. A monsoon variability index, incorporating six key parameters, identifies heightened vulnerability in Vidarbha and Marathwada districts. Structural equation modeling establishes connections between the index, average yield, and cropped area, aiding in the determination of optimal cropping patterns for vulnerable districts. The study proposes a district-level empirical model for monsoon variability, offering practical insights for regional climate action plans and guiding agricultural policies and climate adaptation measures. The below Table 1 gives detailed literature study including the details of the dataset used, comparative analysis of various models used with their performance.

CHAPTER 3

PROPOSED SYSTEM

3. PROPOSED SYSTEM

In this project, a predictive model is designed to recommend suitable crops for cultivation based on various environmental and soil parameters. The proposed system involves a step-by-step process starting from data collection, preprocessing, model building, and evaluation. The goal is to accurately predict the most suitable crop for a given set of conditions, which can aid farmers in making informed decisions.

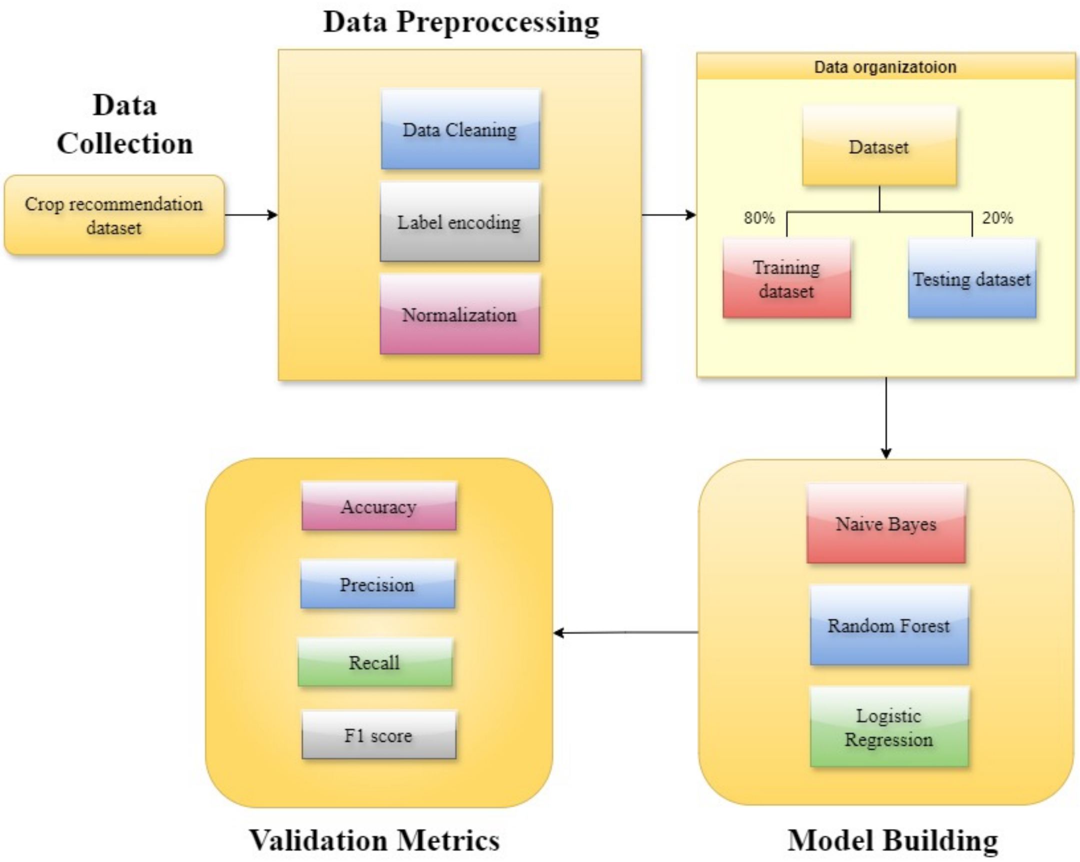

Figure 1.Proposed model architecture

This Figure-1 represents the process of building a machine learning model for crop recommendation. It begins with data collection and preprocessing, which includes steps like data cleaning, label encoding, and normalization. The dataset is then split into training and testing sets, followed by model building using algorithms like Naive Bayes, Random Forest, and Logistic Regression. Finally, the models are evaluated based on validation metrics such as accuracy, precision, recall, and F1 score.

3.1 Input Dataset

The dataset used for building the predictive model consists of a collection of environmental and soil data points. Each data point contains values of various parameters essential in recommending the right crop for the specific conditions.

3.1.1 Detailed Features of the Dataset

The dataset features include:

- **N (Nitrogen content in soil)**: Represents the amount of nitrogen available for the crop.
- **P (Phosphorus content in soil)**: Represents the amount of phosphorus available for the crop.
- **K (Potassium content in soil)**: Represents the amount of potassium available for the crop.
- **Temperature (°C)**: The atmospheric temperature at the location.
- **Humidity (%)**: The level of moisture in the air.
- **pH**: Soil acidity or alkalinity level, which affects nutrient availability.
- **Rainfall (mm)**: The amount of rainfall received.
- **Label**: The target variable, representing the crop type recommended for the given conditions.

Each of these features is directly related to the suitability of crops and is key to building an accurate predictive model.

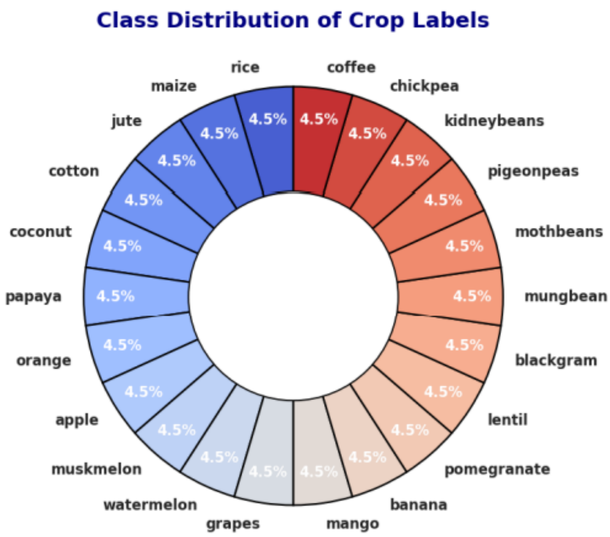

Figure 2.Crop recommendation dataset class distribution of Crop Labels

The figure-2 shows a circular bar chart representing the class distribution of crop labels. Each crop category, including rice, coffee, maize, banana, and others, has an equal distribution of 4.5%, indicating a balanced dataset for classification.

3.2 Data Pre-processing

Before building the model, the dataset undergoes various pre-processing steps to ensure that it is in a suitable form for training. This process includes handling missing data, encoding categorical variables, and feature scaling.

3.2.1 Missing Values

Missing data can severely affect the performance of machine learning models. To ensure that the dataset is complete and ready for training, missing values are identified and handled using appropriate techniques. In this case, any missing values are filled using the mean of the respective columns to maintain the consistency of the dataset.

3.2.2 Feature Scaling

The Min-Max Scaler was applied to normalize the data, scaling all feature values between 0 and 1. This ensures that no feature with larger numerical ranges disproportionately influences the model.

3.2.3 Label Encoding:

By giving each category a unique number, label encoding transforms categorical data input into a format that machine learning models can understand. The algorithm is able to handle non-numerical data in this fashion. Because categories are handled independently, they do not inherently imply an ordinal relationship between them. In order to enable the model to operate with categorical features while preserving the true meaning of the data for accurate learning and prediction, the labels are converted to numeric values

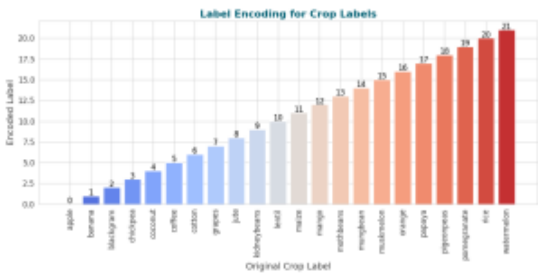

Figure 3. Label Encoding

The figure-3 presents a bar graph showing the label encoding for various crop labels, mapping each crop to a unique numerical value. The crops are ordered by their encoded labels, ranging from apple (encoded as 0) to watermelon (encoded as 21).

3.3 Model Building

After the dataset is pre-processed, various machine learning models are trained to predict the target variable, which is the calories burned. The models used include:

- **Navie bayes**
- **Logistic Regression**
- **Decision Tree**
- **Random Forest**
- **Gradient Boosting**
- **K-Nearest Neighbour's**
- **Support Vector Machine**

The objective is to compare the performance of these models and select the one that provides the most accurate predictions.

3.4 Methodology of the System

The overall methodology of the system is as follows:

1. **Data Loading**: The dataset is loaded into the system for analysis.
2. **Data Pre-processing**: Missing values are handled, and features are scaled.
3. **Feature Selection**: Relevant features (N, P, K, Temperature, Humidity, pH, Rainfall) are identified as critical for predicting the suitable crop.
4. **Model Building**: Multiple models are trained on the dataset, including Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, etc.
5. **Model Evaluation**: Each model is evaluated using various metrics such as accuracy, precision, recall, and F1-score.
6. **Model Selection**: The best-performing model is chosen based on its evaluation metrics.
7. **Prediction**: The selected model is used to make predictions on new data.

This methodology ensures that the system is robust and capable of providing accurate crop recommendations based on environmental and soil parameters.

3.5 Model Evaluation

The trained models are evaluated based on several metrics, including:

- **Accuracy**: Measures the percentage of correctly predicted crops.
- **Precision**: The ratio of correctly predicted crops to the total predicted crops.
- **Recall**: The ratio of correctly predicted crops to the total actual crops.

- **F1-Score**: The harmonic mean of precision and recall, providing a balance between the two.

3.6 Constraints

The system is subject to the following constraints:

- **Dataset Quality**: The accuracy of the model is dependent on the quality of the input data. Incomplete or biased data can result in lower prediction accuracy.
- **Computational Power**: Training complex models like Random Forest and XGBoost may require substantial computational resources, especially for large datasets.
- **Overfitting**: Some models, particularly decision trees, are prone to overfitting if not properly regularized, leading to reduced performance on unseen data.
- **Interpretability**: While tree-based models are interpretable, some complex models can be considered "black boxes," making it challenging to understand how the model arrived at a particular prediction.

3.7 Cost and Sustainability Impact

From a sustainability perspective, machine learning models like Random Forest and XGBoost can be computationally expensive to train, especially on large datasets. High energy consumption is associated with intensive training processes, particularly when using complex models. However, once trained, the models can provide real-time predictions, reducing the need for continuous retraining and thus mitigating energy usage.

Cost-wise, cloud-based platforms offering scalable computational resources, like AWS and Google Cloud, can help manage the training process efficiently. Balancing accuracy with computational cost is crucial, particularly for models that will be deployed on a larger scale or integrated into agricultural advisory systems.

CHAPTER 4

Implementation

4. Implementation

The implementation phase covers the practical application of the proposed crop recommendation system, including setting up the environment, processing the data, and executing the models. The following sections detail the steps required for implementing the crop recommendation model using machine learning.

4.1 Environment Setup

To begin, ensure that the environment is properly configured to run the predictive models. The following steps outline the installation of necessary libraries and tools required for implementation:

1. **Programming Language**: The implementation is carried out using Python, a popular language for machine learning.
2. **Libraries**:
   - **Pandas**: For data manipulation and preprocessing.
   - **NumPy**: For numerical computations.
   - **Scikit-learn**: For implementing machine learning models.
   - **Matplotlib/Seaborn**: For visualizing the results.
   - **XGBoost**: For implementing XGBoost model.
3. **Installation**: Install the required libraries using pip:

```
pip install pandas numpy scikit-learn matplotlib seaborn xgboost
```

4. **Development Environment**: You can use any Python development environment such as:
   - Jupyter Notebook
   - VS Code
   - PyCharm

4.2 Sample Code for Preprocessing and Model Operations

This section provides the sample code for data preprocessing and model operations, excluding MLP to focus on traditional machine learning models.

1. **Data Preprocessing**:

Load the Dataset:

```
import pandas as pd
data = pd.read_csv('crop_recommendation_data.csv')
```

Remove Duplicate Entries:

```
data.drop_duplicates(inplace=True)
```

Data Splitting:

```
from sklearn.model_selection import train_test_split
```

# Split the data into training (80%) and testing (20%) sets
X = data[['N', 'P', 'K', 'temperature', 'humidity', 'ph', 'rainfall']]
y = data['label']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

Feature Scaling (Min-Max Normalization):

```
from sklearn.preprocessing import MinMaxScaler
# Apply Min-Max Scaling to standardize the range
scaler = MinMaxScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```

2. **Model Building and Training**: The following is a sample of how to implement and train different machine learning models for predicting calories burned.

Naive Bayes:

```
from sklearn.naive_bayes import GaussianNB
# Initialize and train the Naive Bayes model
nb_model = GaussianNB()
nb_model.fit(X_train, y_train)
```

Logistic Regression:

```
from sklearn.linear_model import LogisticRegression
# Initialize and train the Logistic Regression model
lr_model = LogisticRegression()
lr_model.fit(X_train, y_train)
```

Random Forest:

```
from sklearn.ensemble import RandomForestClassifier
# Initialize and train the Random Forest model
rf_model = RandomForestClassifier(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)
```

Gradient Boosting:

```
from sklearn.ensemble import GradientBoostingClassifier
# Initialize and train the Gradient Boosting model
gb_model = GradientBoostingClassifier(n_estimators=100, learning_rate=0.1, random_state=42)
gb_model.fit(X_train, y_train)
```

- **Model Evaluation**: Once the models are trained, evaluate their performance using metrics such as accuracy, precision, recall, and F1-score.

  Evaluate Models:

```
from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score

def evaluate_model(model, X_test, y_test):
    y_pred = model.predict(X_test)
    print(f"Accuracy: {accuracy_score(y_test, y_pred)}")
    print(f"Precision: {precision_score(y_test, y_pred, average='weighted')}")
    print(f"Recall: {recall_score(y_test, y_pred, average='weighted')}")
    print(f"F1-Score: {f1_score(y_test, y_pred, average='weighted')}")

# Evaluate Naive Bayes model
print("Naive Bayes Performance:")
evaluate_model(nb_model, X_test, y_test)

# Evaluate Logistic Regression model
print("Logistic Regression Performance:")
evaluate_model(lr_model, X_test, y_test)

# Evaluate Random Forest model
print("Random Forest Performance:")
evaluate_model(rf_model, X_test, y_test)

# Evaluate Gradient Boosting model
print("Gradient Boosting Performance:")
evaluate_model(gb_model, X_test, y_test)
```

3. **Model Selection and Prediction**: After evaluating the models, choose the one with the best performance metrics and use it for recommending crops on new data.

- **Prediction Example**:

```
# Predict calories burned using the best model
# Predict the best crop using the best-performing model
best_model = rf_model  # Assuming Random Forest performed the best
new_data = [[90, 42, 43, 25.0, 80, 6.5, 200]]  # Example input: N, P, K, temp, humidity, pH, rainfall
scaled_data = scaler.transform(new_data)
predicted_crop = best_model.predict(scaled_data)
```

```
print(f"Recommended Crop: {predicted_crop[0]}")
```

Summary of Implementation

The implementation process is structured to ensure efficient data preprocessing and model building using several popular machine learning algorithms. The focus is on removing duplicates, feature scaling, and training various models like Naive Bayes, Logistic Regression, Random Forest, and Gradient Boosting. Each model is evaluated for performance, and the best model is selected for making crop recommendations.

CHAPTER 5

Experimentation and Result Analysis

5. Experimentation and Result Analysis

In this section, we analyze the performance of various machine learning models used for crop recommendation based on climate and soil data. The primary goal was to identify the most effective model by comparing key performance metrics.

Experimentation Setup

The experiment was structured to evaluate different models by splitting the dataset into training (80%) and testing (20%) sets. Each model was trained on the training data and evaluated on the test set using the following metrics:

- **Accuracy**
- **Precision**
- **Recall**
- **F1-Score**

The models evaluated include:

- Naive Bayes
- Random Forest
- Bagging
- Gradient Boosting
- k-Nearest Neighbors (KNN)
- Support Vector Machine (SVM)
- Logistic Regression

Result Analysis

The following insights were drawn from the results:

- **Naive Bayes** emerged as the best-performing model with an **Accuracy** of 99.55%, **Precision** of 99.63%, **Recall** of 99.55%, and **F1-Score** of 99.54%. Its probabilistic approach enabled it to handle feature interactions effectively, leading to highly accurate predictions. This model's strong performance across all metrics makes it the most reliable choice for crop recommendation.

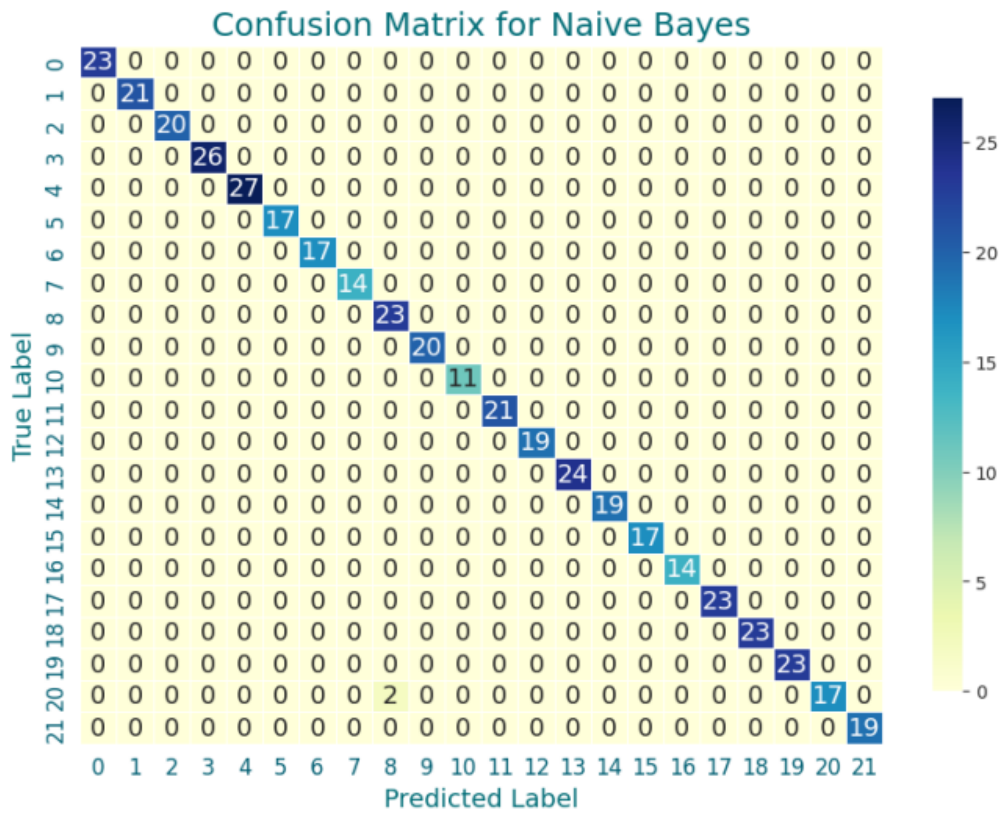

Figure 5. Confusion matrix of Naïve Bayes

The figure-5 shows a confusion matrix for Naive Bayes classification, with perfect diagonal values indicating accurate predictions. The matrix is color-coded, where darker shades represent higher values, showing minimal misclassification across labels.

- **Random Forest** also performed well, with an **Accuracy** of 99.32%, **Precision** of 99.37%, **Recall** of 99.32%, and **F1-Score** of 99.32%. Its ability to aggregate multiple decision trees helped capture complex patterns in the data. Although slightly less accurate than Naive Bayes, it still showed robust performance, making it a suitable alternative.
- **Bagging** achieved an **Accuracy** of 98.64%, and its **Precision**, **Recall**, and **F1-Score** were all around 98.67%. This indicates good performance, though not as high as Naive Bayes and Random Forest. Bagging was effective at reducing overfitting but did not reach the top accuracy levels.

- **Gradient Boosting** performed commendably with an **Accuracy** of 98.18%, **Precision** of 98.43%, **Recall** of 98.18%, and **F1-Score** of 98.19%. It managed to capture non-linear relationships well but had slightly higher error rates than Random Forest and Naive Bayes, suggesting it may need more hyperparameter tuning.

- **k-Nearest Neighbors (KNN)** and **Support Vector Machine (SVM)** both achieved **Accuracy** scores of 96.82%, with KNN showing a **Precision** of 97.19% and SVM achieving 97.30%. These models demonstrated solid performance, but the increased computational requirements, especially for SVM, could make them less practical for real-time deployment.

- **Logistic Regression** had the lowest performance metrics, with an **Accuracy** of 91.82%, **Precision** of 93.44%, **Recall** of 91.82%, and **F1-Score** of 91.72%. While useful for simpler datasets, it struggled to handle the complexity of the features, leading to lower predictive power.

Table 2: Metrics of Performance for the Suggested Models

Model	Accuracy	Precision	Recall	F1-Score
Naive Bayes	0.9955	0.9963	0.9955	0.9954
Random Forest	0.9932	0.9937	0.9932	0.9932
Bagging	0.9864	0.9867	0.9864	0.9864
Gradient Boosting	0.9818	0.9843	0.9818	0.9819
k-Nearest Neighbours	0.9682	0.9719	0.9682	0.9682
support vector machine	0.9682	0.9730	0.9682	0.9682
Logistic Regression	0.9182	0.9344	0.9182	0.9172

The table-2 compares performance metrics (Accuracy, Precision, Recall, and F1-Score) for several machine learning models. Naive Bayes achieves the highest accuracy (0.9955) and F1-score (0.9954), while Logistic Regression shows the lowest values across all metrics.

Summary

The experiment highlights the importance of model selection based on accuracy, precision, and computational efficiency. **Naive Bayes** stood out as the top performer across all metrics, offering a combination of high accuracy and efficient computation, making it ideal for real-time crop recommendation systems. **Random Forest** was also a strong contender, offering slightly lower accuracy but robust overall performance.

Visual Representation of Results

The comparison of the models' metrics can be summarized as follows:

- **Top Performer**: Naive Bayes (Accuracy: 99.55%)
- **Strong Alternative**: Random Forest (Accuracy: 99.32%)
- **Effective but Less Accurate**: Bagging, Gradient Boosting
- **Moderate Performance**: KNN, SVM
- **Least Effective**: Logistic Regression

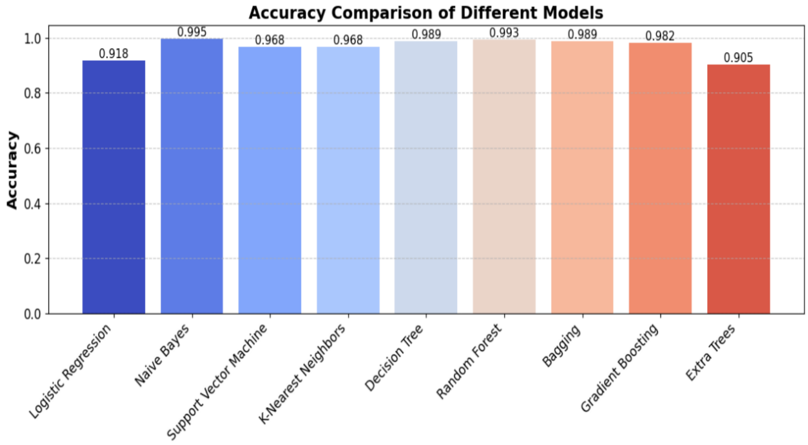

Figure 4. Accuracy Comparison of Different Models

The performance differences emphasize how different models can be suitable for different use cases, but for this particular task, **Naive Bayes** provided the best balance of accuracy and practicality.

Table 3: Comparison with previous model

Metrics	Existing model [1]	Proposed model
Accuracy	0.9932	0.9955
Precision	0.9950	0.9963
Recall	0.9963	0.9955
F1-score	0.9959	0.9954

Table 3 shows the comparison of the suggested and current models. Both models were excellent; however, the accuracy and precision of the suggested model are improved to around 99.55% and 99.63%, respectively, whereas the accuracy and precision of the existing model are 99.50% and 99.32%, respectively. For both models, the F1-score and Recall values are nearly identical

CHAPTER-6

Conclusion

6. Conclusion

This project effectively demonstrated the application of machine learning models for developing a climate-based crop recommendation system. By leveraging climate and soil data, we evaluated several models, including Naive Bayes, Random Forest, Bagging, Gradient Boosting, and others, to determine the most accurate and reliable approach for predicting the best crops for specific regions.

Through rigorous experimentation and analysis, **Naive Bayes** emerged as the most effective model, delivering the highest accuracy and consistency across various performance metrics, including precision, recall, and F1-score. The model's ability to handle probabilistic relationships between features allowed it to excel in predicting optimal crop choices, making it well-suited for real-world agricultural applications.

The experimentation process underscored the importance of model selection and the role of evaluation metrics in identifying the best-performing algorithms. While models like Logistic Regression provided a baseline understanding, more sophisticated approaches like Naive Bayes and Random Forest proved essential in capturing the complex relationships within the dataset, leading to higher accuracy and better generalization.

Additionally, the project highlighted the critical role of data preprocessing, including handling missing values, scaling, and feature selection, in enhancing the performance of machine learning models. These steps ensured that the algorithms could effectively learn from the data and produce accurate predictions.

Future Work could involve expanding the dataset to include additional environmental factors or agricultural features, such as soil pH, precipitation levels, and crop rotation patterns, to further refine the recommendations. Additionally, fine-tuning the hyperparameters of the top-performing models may lead to even greater accuracy and robustness, especially when deploying the system on a larger scale.

Ultimately, this study contributes to the ongoing advancement of predictive analytics in agriculture, providing farmers and agronomists with valuable insights that can help optimize crop selection and improve yield outcomes. By applying machine learning to agriculture, we can empower decision-makers with data-driven guidance, leading to more resilient and sustainable farming practices

REFERENCES

1. Pachade, R. S., & Sharma, A. (2022). Machine Learning for WeatherSpecific Crop Recommendation. In International Journal of Health Sciences, 6(S8), 4527–4537. doi: 10.53730/ijhs.v6nS8.13222.
2. Shams, M.Y., Gamel, S.A., & Talaat, F.M. (2024). Enhancing crop recommendation systems with explainable artificial intelligence: a study on agricultural decision-making. Neural Computing and Applications, 36(5695–5714). doi: 10.1007/s00521-023-09391-2.
3. Mahale, Y., Khan, N., Kulkarni, K., et al. (2024). Crop recommendation and forecasting system for Maharashtra using machine learning with LSTM: a novel expectation-maximization technique. Discover Sustainability, 5(134). doi: 10.1007/s43621-024-00292-5.
4. 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, 7(3), 558-569.
5. Pande, S. M., Ramesh, P. K., Anmol, A., Aishwarya, B. R., Rohilla, K., & Shaurya, K. (2021). Crop recommender system using machine learning approach. 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), 1066-1071. doi: 10.1109/ICCMC51019.2021.9418351.
6. Ingle, A. (2020). Crop Recommendation Dataset. Retrieved from https://www.kaggle.com/datasets/atharvaingle/crop-recommendationdataset
7. Durai SKS, Shamili MD. Smart farming using machine learning and deep learning techniques. Decis Anal J. 2022;3: 100041.
8. Sharma P, Dadheech P, Senthil ASK. Ai-enabled crop recommendation system based on soil and weather patterns. In: Artifcial Intelligence Tools and Technologies for Smart Farming and Agriculture Practices, pp. 184–199. IGI Global, 2023.
9. Elavarasan D, Vincent PD. Crop yield prediction using deep reinforcement learning model for sustainable Agrarian applications. IEEE Access. 2020;8:86886–901.

10. Cai Y, Guan K, Peng J, Wang S, Seifert C, Wardlow B, Li Z. A high-performance and in-season classifcation system of feld-level crop types using time-series Landsat data and a machine learning approach. Remote Sens Environ. 2018;210:35–47.
11. Ed-daoudi R, Alaoui A, Ettaki B, Zerouaoui J. A predictive approach to improving agricultural productivity in morocco through crop recommendations. Int J Adv Comput Sci Appl 2023;14(3).
12. Moon MH, Marjan MA, Uddin MP, Ibn Afal M, Kadry S, Ma S, Nam Y. Ensemble machine learning-based recommendation system for efective prediction of suitable agricultural crop cultivation. Front Plant Sci. 2023;14:1234555.
13. Nayak HS, Silva JV, Parihar CM, Krupnik TJ, Sena DR, Kakraliya SK, Jat HS, Sidhu HS, Sharma PC, Jat ML, et al. Interpretable machine learning methods to explain on-farm yield variability of high productivity wheat in northwest India. Field Crops Res. 2022;287: 108640.
14. Van Klompenburg T, Kassahun A, Catal C. Crop yield prediction using machine learning: a systematic literature review. Comput Electron Agric. 2020;177: 105709.
15. Banerjee S, Mondal AC. A region-wise weather data-based crop recommendation system using diferent machine learning algorithms. Int J Intell Syst Appl Eng. 2023;11(3):283–97.
16. Su Y-x, Xu H, Yan L-j. Support vector machine-based open crop model (sbocm): case of rice production in china. Saudi J Biol Sci. 2017;24(3):537–47.
17. Kedlaya A, Sana A, Bhat BA, Kumar S, Bhat N, et al. An efcient algorithm for predicting crop using historical data and pattern matching technique. Global Transit Proc. 2021;2(2):294–8.
18. Samuel P, Sahithi B, Saheli T, Ramanika D, Kumar NA. Crop price prediction system using machine learning algorithms. Quest J Softw Eng Simul. 2020.
19. Gupta T, Maggu S, Kapoor B. Crop prediction using machine learning. 2023.
20. Chandana C, Parthasarathy G. Efcient machine learning regression algorithm using naïve Bayes classifer for crop yield prediction and optimal utilization of fertilizer. Int J Performabil Eng. 2022;18(1).
21. Todmal RS. Future climate change scenario over Maharashtra, western India: implications of the regional climate model (remo-2009) for the understanding of agricultural vulnerability. Pure Appl Geophys. 2021;178(1):155–68.

22. Cedric LS, Adoni WYH, Aworka R, Zoueu JT, Mutombo FK, Krichen M, Kimpolo CLM. Crops yield prediction based on machine learning models: case of west African countries. Smart Agric Technol. 2022;2: 100049.
23. Alebele Y, Wang W, Yu W, Zhang X, Yao X, Tian Y, Zhu Y, Cao W, Cheng T. Estimation of crop yield from combined optical and sar imagery using Gaussian kernel regression. IEEE J Sel Topics Appl Earth Observ Remote Sens. 2021;14:10520–34.
24. Nti IK, Zaman A, Nyarko-Boateng O, Adekoya AF, Keyeremeh F. A predictive analytics model for crop suitability and productivity with tree-based ensemble learning. Decis Anal J. 2023;8: 100311.
25. Liu J, Yang K, Tariq A, Lu L, Soufan W, El Sabagh A. Interaction of climate, topography and soil properties with cropland and cropping pattern using remote sensing data and machine learning methods. Egypt J Remote Sens Space Sci. 2023;26(3):415–26.
26. Johnston DB, Pembleton KG, Huth NI, Deo RC. Comparison of machine learning methods emulating process driven crop models. Environ Modell Softw. 2023;162: 105634.
27. Raja S, Sawicka B, Stamenkovic Z, Mariammal G. Crop prediction based on characteristics of the agricultural environment using various feature selection techniques and classifers. IEEE Access. 2022;10:23625–41.
28. Ghadge R, Kulkarni J, More P, Nene S, Priya R. Prediction of crop yield using machine learning. Int Res J Eng Technol (IRJET). 2018;5:2237–9.
29. Devan K, Swetha B, Sruthi PU, Varshini S. Crop yield prediction and fertilizer recommendation system using hybrid machine learning algorithms. In: 2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT), pp. 171–175, 202