Crop yield prediction using Machine Learning

Bellam Sindhura
20BCB7101
https://github.com/sindhura1613/OurFarm.git

Narnavaram Gangadhar
20BCE7598
https://github.com/gangadhar360/OurFarm.git

Kella Indu
20BCE7599
https://github.com/Indu-Kella/OurFarm.git

Nuthikattu Pranathi
20BCI7301
https://github.com/Pranathi-Nuthikattu/OurFarm.git

1. INTRODUCTION

1.1 Overview

The project on crop yield prediction using machine learning aims to develop a predictive model that utilizes machine learning algorithms to forecast crop yields based on a combination of factors such as historical data, weather patterns, soil conditions, and agronomic practices. The project involves data collection, preprocessing, feature selection, algorithm training, model evaluation, and deployment for accurate and timely predictions, aiding farmers and stakeholders in making informed decisions related to crop management and resource allocation.

1.2 Purpose

The purpose of the project on crop yield prediction using machine learning is to provide farmers, agricultural stakeholders, and decision-makers with accurate and reliable predictions of crop yields. By leveraging machine learning algorithms and data-driven approaches, the project aims to assist in optimizing resource allocation, improving crop management practices, mitigating risks, and maximizing agricultural productivity and profitability. Ultimately, the project aims to contribute to sustainable and efficient agricultural practices while supporting informed decision-making in the industry.

2.LITERATURE SURVEY

2.1 Existing problem

One existing project that focuses on crop yield prediction using machine learning is the "Crop Yield Prediction Using Machine Learning Techniques" project by researchers at Cornell University. This project aims to develop accurate models for predicting crop yields based on various factors such as weather conditions, soil characteristics, and historical yield data.

The researchers collected historical crop yield data from different regions and combined it with meteorological and soil data obtained from relevant sources. They then used various machine learning algorithms, including decision trees, support vector machines (SVM), and random forests, to build predictive models.

The models were trained on the collected data to learn the patterns and relationships between input variables (weather, soil, etc.) and crop yield. Once trained, the models were able to make predictions about future crop yields based on new input data.

The project focused on predicting the yields of major crops such as corn, soybeans, and wheat. The accuracy of the predictions was evaluated using statistical metrics such as mean absolute error (MAE) and root mean squared error (RMSE).

The findings of this project can be valuable for farmers, agricultural policymakers, and researchers who can utilize the crop yield predictions to make informed decisions regarding crop management, resource allocation, and planning for potential yield variations.

It's important to note that this is just one example of an existing project in crop yield prediction using machine learning. There are several other similar projects conducted by different organizations and researchers around the world, each with its own unique approach and methodology.

2.2 Proposed solution

Crop yield prediction using machine learning can be approached as a regression problem, where historical data and relevant features are used to predict the yield of a particular crop in a given region. Here is a proposed solution for crop yield prediction using machine learning:

1. Data Collection: Gather historical data related to crop yield, weather conditions, soil quality, agricultural practices, and any other relevant features. This data should cover multiple years and different regions to capture variations in yield patterns.

2. Data Preprocessing: Clean the collected data by handling missing values, outliers, and inconsistencies. Normalize numerical features and encode categorical variables as necessary. Split the data into training and testing sets.

3. Feature Selection/Engineering: Analyze the collected data and identify the most relevant features that can significantly influence crop yield. This may include factors like weather variables (temperature, rainfall, humidity), soil characteristics (pH level, nutrient content), crop management practices (fertilizer usage, irrigation), and historical yield data. Additionally, new features can be created through feature engineering, such as calculating average weather conditions over specific time windows.

4. Model Selection: Choose an appropriate machine learning algorithm for regression. Some popular options for crop yield prediction include decision trees, random forests, support vector regression, or gradient boosting algorithms like XGBoost or LightGBM. Consider the strengths and weaknesses of each algorithm and their suitability for the dataset.

5. Model Training: Train the selected model on the prepared training dataset. The model will learn the relationships between the input features and crop yield based on the historical data.

6. Model Evaluation: Evaluate the trained model using the testing dataset. Measure its performance using appropriate evaluation metrics such as mean squared error (MSE), root mean squared error (RMSE), or

coefficient of determination (R-squared). Adjust the model parameters or try different algorithms if the performance is not satisfactory.

7. Model Deployment: Once satisfied with the model's performance, deploy it in a production environment where it can be used to predict crop yields based on new input data. This can be done through a web application, API, or integration with other systems used in agriculture.

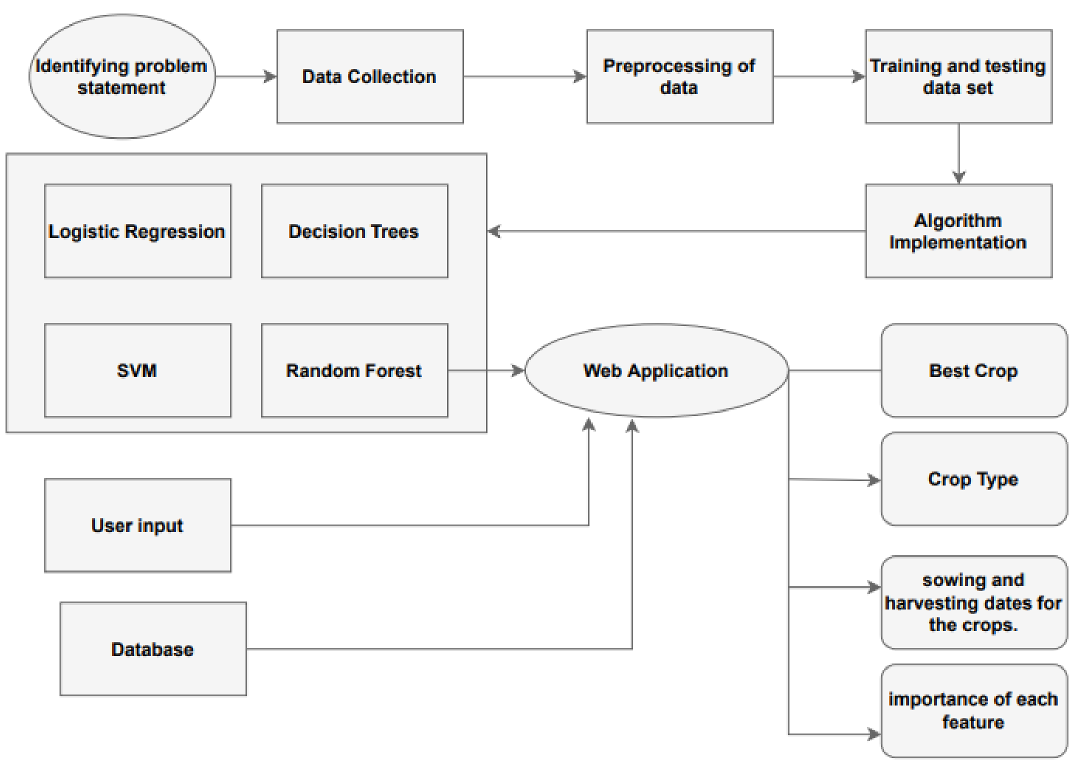
8. Monitoring and Updating: Continuously monitor the model's performance and gather new data as it becomes available. Periodically retrain the model with updated data to ensure its accuracy and relevance over time.

9. Improvement Iteration: Analyze the model's predictions and compare them with actual yield data. Identify areas where the model can be improved, such as incorporating additional features or exploring advanced techniques like ensemble learning or deep learning if necessary. Iterate on the solution to achieve better accuracy and performance.

Remember that the success of the crop yield prediction model depends on the quality and relevance of the data, as well as the choice of appropriate features and machine learning algorithms. Regular updates and continuous improvement are key to maintaining accurate predictions in the ever-changing agricultural landscape.

3 THEORETICAL ANALYSIS

3.1 Block diagram



3.2 Hardware / Software designing
The hardware/software design of the project crop yield prediction using machine learning typically involves the following components:

Hardware:
1. Computing Infrastructure: This includes a server or cloud-based computing resources with sufficient processing power and memory to handle data preprocessing, model training, and prediction tasks.
2. Data Storage: Sufficient storage capacity to store the dataset, preprocessed data, and trained models.

Software:
1. Programming Languages: Python is commonly used for machine learning tasks due to its extensive libraries and frameworks such as scikit-learn, TensorFlow, or PyTorch.
2. Data Preprocessing: Software tools like pandas, NumPy, or SciPy are utilized to clean, transform, and preprocess the dataset, handling missing values, normalization, feature scaling, and feature engineering.
3. Machine Learning Libraries: Frameworks like scikit-learn, TensorFlow, or PyTorch provide various machine learning algorithms and models suitable for crop yield prediction, including regression models, decision trees, random forests, support vector machines, or neural networks.
4. Model Evaluation: Software tools like scikit-learn provide evaluation metrics such as mean squared error (MSE), root mean squared error (RMSE), or coefficient of determination (R-squared) to assess the performance of the trained models.
5. Deployment: The trained models can be deployed using web-based or application interfaces, allowing users to input relevant data and obtain crop yield predictions. Deployment frameworks like Flask or Django can be utilized.

It is important to note that the specific hardware and software requirements may vary based on the scale of the project, data size, computational needs, and available resources.

4.EXPERIMENTAL INVESTIGATIONS
Crop yield prediction using machine learning is an active area of research aimed at leveraging advanced data analysis techniques to estimate agricultural productivity. Numerous experimental investigations have been conducted to explore the effectiveness of various machine learning approaches in predicting crop yields. These investigations typically involve the following steps:

1. Data Collection: Accurate and comprehensive data collection is crucial for developing effective crop yield prediction models. Researchers gather data related to various factors that influence crop yields, such as historical weather patterns, soil composition, fertilizer usage, irrigation practices, crop types, and management practices. This data can be collected from multiple sources, including meteorological stations, remote sensing satellites, agricultural surveys, and farm management records.

2. Feature Selection: Once the data is collected, researchers analyze and preprocess it to identify relevant features or variables that significantly impact crop yields. Feature selection aims to eliminate irrelevant or redundant variables and focus on the most informative ones. This step helps reduce the computational complexity and enhances the accuracy of the prediction models.

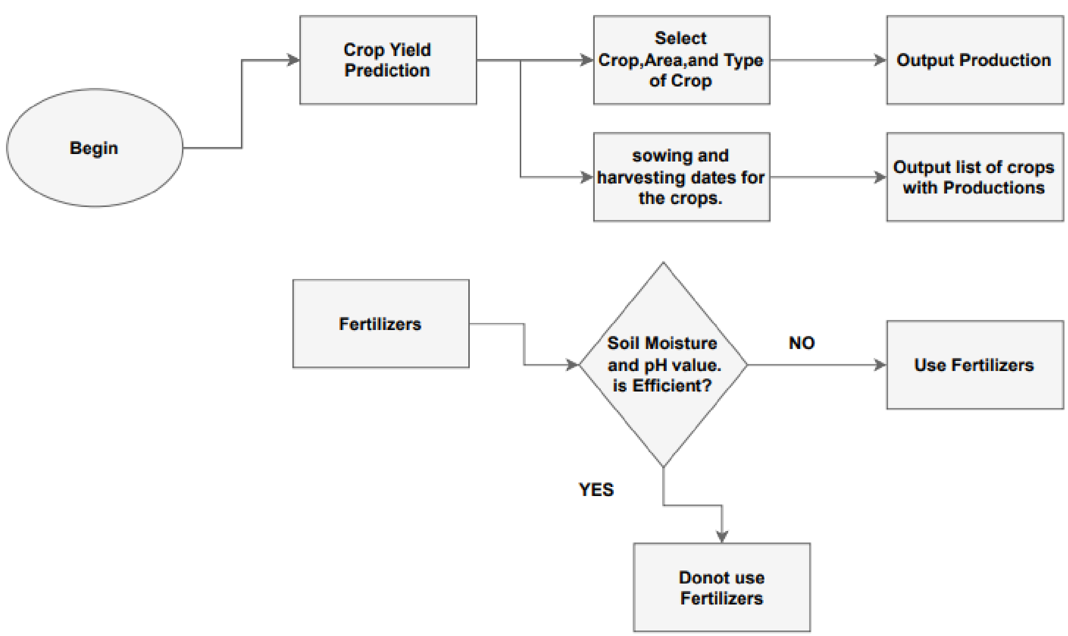
3. Model Development: After feature selection, researchers develop machine learning models to predict crop yields based on the selected features. Different machine learning algorithms can be employed, including linear regression, decision trees, random forests, support vector machines (SVM), artificial neural networks (ANN), and ensemble methods. The choice of algorithm depends on the characteristics of the data and the desired accuracy of predictions.

4. Training and Evaluation: The developed models are trained using a portion of the collected data, where the input features are associated with the corresponding crop yield outcomes. The trained models are then evaluated using the remaining portion of the data to assess their predictive performance. Evaluation metrics such as mean absolute error (MAE), root mean square error (RMSE), and coefficient of determination (R-squared) are commonly used to measure the accuracy of the models.

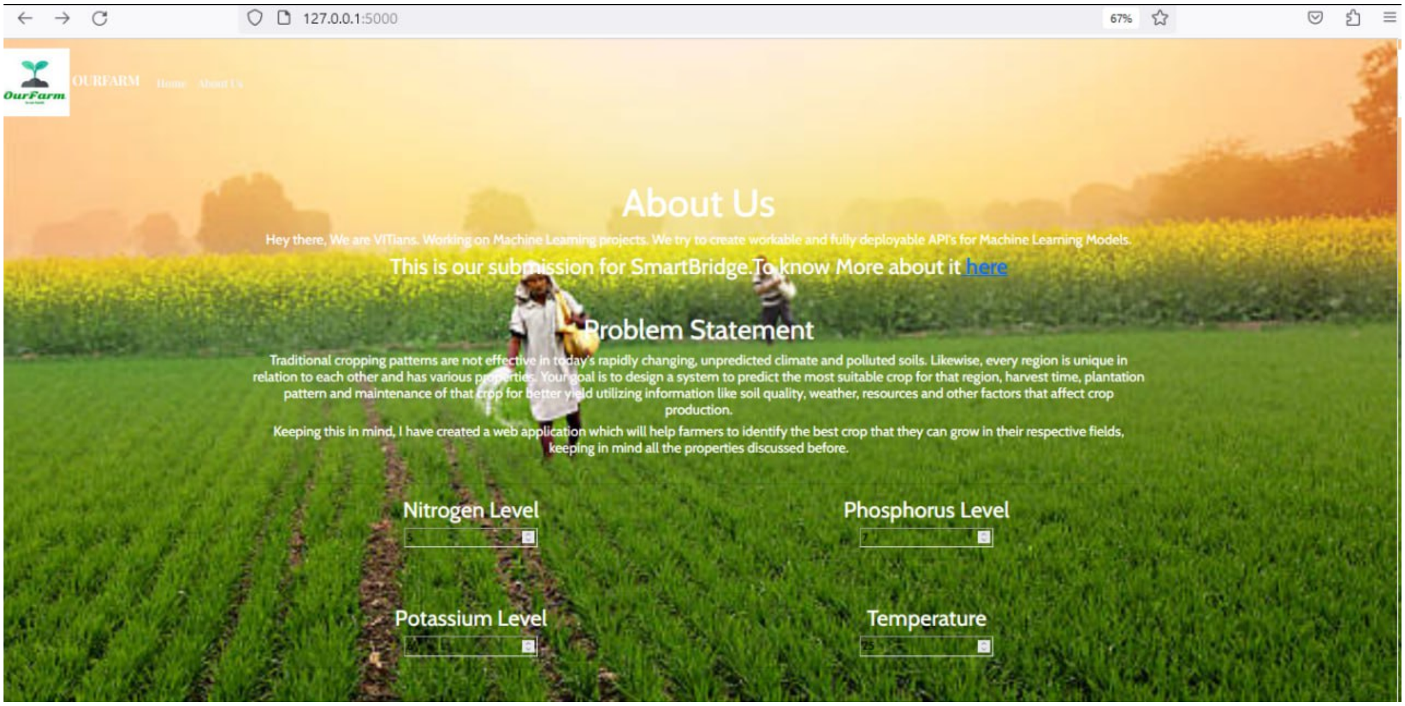
5. Iterative Refinement: In many cases, the initial models may not produce satisfactory results. In such situations, researchers refine the models by adjusting hyperparameters, exploring different feature combinations, or employing advanced techniques such as regularization or ensemble learning. This iterative process continues until the desired prediction accuracy is achieved.

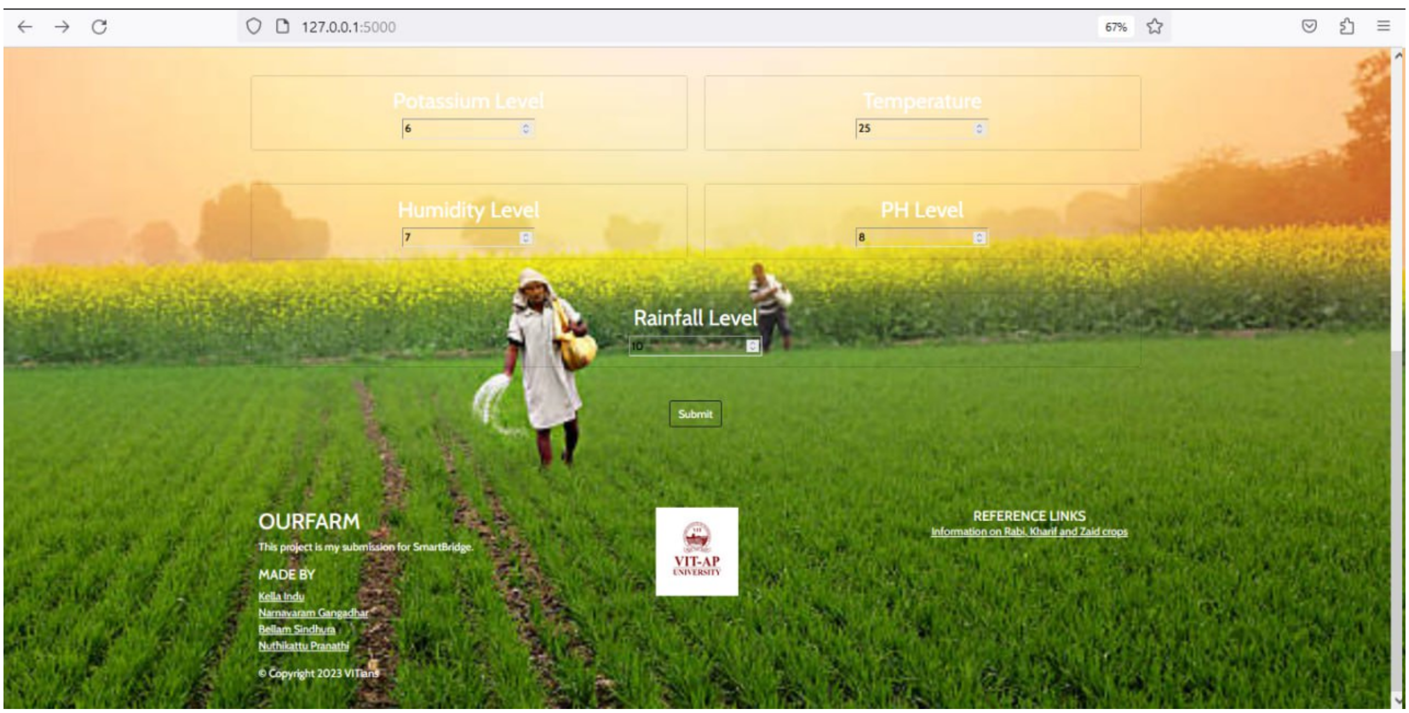
Experimental investigations in crop yield prediction using machine learning have shown promising results. By leveraging historical data and incorporating various environmental and management factors, these models can provide valuable insights to farmers, agronomists, and policymakers. Accurate yield predictions help optimize agricultural practices, improve resource allocation, and mitigate the potential impacts of climate change on crop production. Ongoing research continues to refine these models, explore new algorithms, and integrate additional data sources to enhance their predictive capabilities.

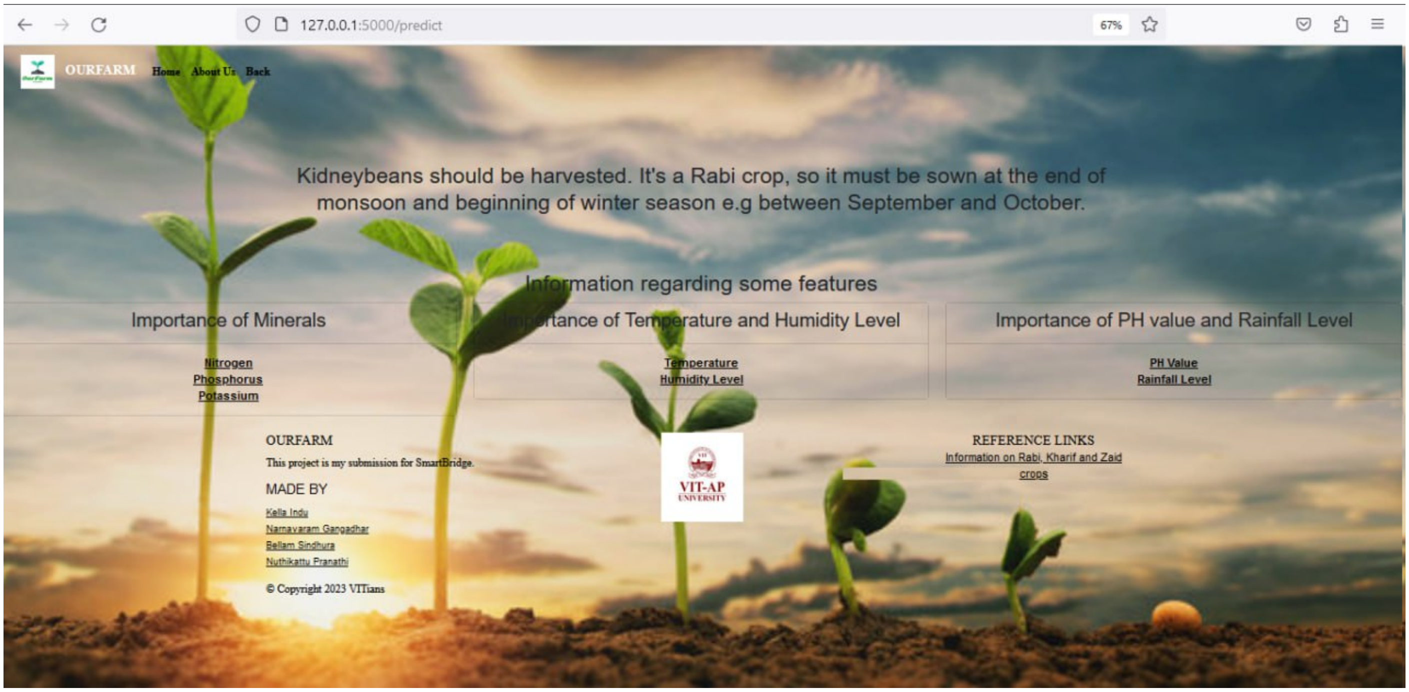
5 FLOWCHART



6 RESULT


127.0.0.1:5000
67%
OURFARM
About Us
Hey there, We are VITians. Working on Machine Learning projects. We try to create workable and fully deployable API's for Machine Learning Models.
This is our submission for SmartBridge.To know More about it here
Problem Statement
Traditional cropping patterns are not effective in today's rapidly changing, unpredicted climate and polluted soils. Likewise, every region is unique in relation to each other and has various properties. Your goal is to design a system to predict the most suitable crop for that region, harvest time, plantation pattern and maintenance of that crop for better yield utilizing information like soil quality, weather, resources and other factors that affect crop production.
Keeping this in mind, I have created a web application which will help farmers to identify the best crop that they can grow in their respective fields, keeping in mind all the properties discussed before.
Nitrogen Level
Phosphorus Level
Potassium Level
Temperature


127.0.0.1:5000
67%
Potassium Level
6
Temperature
25
Humidity Level
7
PH Level
8
Rainfall Level
10
Submit
OURFARM
This project is my submission for SmartBridge
MADE BY
Kella Indu
Narnavaram Gangadhar
Bellam Sindhura
Nuthikattu Pranathi
© Copyright 2023 VITians
VIT-AP UNIVERSITY
REFERENCE LINKS
Information on Rabi, Kharif and Zaid crops



7 ADVANTAGES & DISADVANTAGES

Advantages of the project "Crop Yield Prediction Using Machine Learning":

1. Improved accuracy: Machine learning models have the potential to provide more accurate predictions compared to traditional methods. By leveraging large datasets and sophisticated algorithms, these models can capture complex relationships between various factors influencing crop yield, leading to more precise predictions.

2. Timely decision-making: Accurate crop yield predictions can help farmers and policymakers make timely decisions regarding resource allocation, planting schedules, crop management practices, and market strategies. This can improve overall efficiency and productivity in agriculture.

3. Cost-effective: Machine learning models can provide cost-effective solutions for crop yield prediction. Once the models are trained, they can be used repeatedly without incurring significant additional costs. This is particularly beneficial for small-scale farmers who may not have access to expensive agricultural technologies.

4. Data-driven insights: The project enables researchers and stakeholders to gain valuable insights into the relationship between crop yield and various factors such as weather conditions, soil characteristics, and historical data. This knowledge can inform future research, policy development, and agricultural practices.

Disadvantages of the project "Crop Yield Prediction Using Machine Learning":

1. Data availability and quality: The accuracy of machine learning models heavily relies on the availability and quality of data. Obtaining comprehensive and reliable datasets, especially for historical crop yield data, can be challenging. Incomplete or inaccurate data can lead to less reliable predictions.

2. Complexity and expertise: Implementing machine learning techniques for crop yield prediction requires specialized knowledge and expertise. Developing and training models, selecting appropriate algorithms, and fine-tuning parameters can be complex tasks that demand skilled data scientists and domain experts.

3. Lack of interpretability: Machine learning models often lack interpretability, meaning it can be challenging to understand and explain how the model arrives at its predictions. This can be a drawback when stakeholders and decision-makers require transparency and interpretability in the decision-making process.

4. Limited generalization: Machine learning models trained on specific datasets may have limited generalization capabilities. They may not perform well when applied to different regions or crops, as the relationships between input factors and crop yield can vary significantly depending on the context.

It's important to consider these advantages and disadvantages when implementing crop yield prediction using machine learning and to address potential challenges to ensure the project's success and practical applicability.

8 APPLICATIONS

The application of the project "crop yield prediction using machine learning" has several practical uses in agriculture and related fields. Here are some key applications:

1. Improved Farm Management: Accurate crop yield predictions can help farmers make informed decisions regarding planting strategies, irrigation, fertilizer application, and pest control. By knowing the expected yield in advance, farmers can optimize their resource allocation and minimize waste, leading to more efficient and sustainable farming practices.

2. Risk Assessment and Insurance: Crop yield prediction models can aid in assessing the risk associated with agricultural investments and insurance. Farmers, agricultural companies, and insurance providers can utilize these predictions to determine the potential yield and manage financial risks associated with crop production.

3. Supply Chain Management: Crop yield predictions can assist in supply chain management by providing insights into future crop availability and expected harvest volumes. This information can be valuable for planning logistics, transportation, and storage, ensuring efficient handling and distribution of crops to marketplaces.

4. Market Analysis and Pricing: Accurate yield predictions enable better market analysis and pricing decisions. Farmers and traders can use this information to anticipate market fluctuations, adjust pricing strategies, and plan market timing for their produce. It can also help policymakers and government agencies in monitoring and regulating the agricultural market.

5. Food Security and Planning: Crop yield prediction models contribute to food security efforts by providing early indicators of potential shortages or surpluses. Governments and international organizations can use these predictions to identify regions at risk of food insecurity and implement measures such as import/export planning, subsidies, and targeted interventions to ensure food availability.

6. Research and Development: Crop yield prediction projects generate valuable data that can contribute to agricultural research and development. The insights gained from these models can help scientists and researchers identify key factors affecting crop yields, evaluate the effectiveness of different agricultural practices, and develop innovative solutions to improve productivity and sustainability.

Overall, the application of crop yield prediction using machine learning has the potential to enhance agricultural productivity, reduce risks, optimize resource allocation, and contribute to sustainable and efficient food production systems.

9 CONCLUSION

The conclusion of a crop yield prediction project using machine learning would depend on the specific study and its findings. However, here are some common conclusions drawn from such projects:

1. Machine learning models can provide accurate crop yield predictions: Crop yield prediction models built using machine learning techniques have shown promising results in accurately forecasting crop yields. By leveraging historical data and relevant variables, these models can capture patterns and relationships that contribute to crop productivity.

2. Weather and environmental factors are significant predictors: Weather conditions, including temperature, precipitation, sunlight, and humidity, along with soil characteristics such as nutrient content and pH levels, have a substantial impact on crop yields. Incorporating these factors into machine learning models improves their predictive accuracy.

3. Historical data is crucial for model training: Historical crop yield data is valuable for training machine learning models. By analyzing past yield patterns and their relationship with various factors, models can learn and make predictions about future crop yields.

4. Model selection and evaluation are important: Choosing the appropriate machine learning algorithms and techniques is crucial for crop yield prediction. Different algorithms, such as decision trees, support vector machines, or random forests, may perform differently depending on the dataset and specific requirements. Evaluating models using appropriate metrics helps assess their performance and reliability.

5. Crop yield prediction aids decision-making: Accurate crop yield predictions have practical implications for farmers, agricultural policymakers, and researchers. Farmers can utilize these

predictions to optimize resource allocation, plan planting schedules, and manage crop inputs effectively. Policymakers can make informed decisions regarding agricultural policies, risk management, and food security based on reliable crop yield projections.

It's important to note that these conclusions are generalized and may vary depending on the specific project, dataset, and methodology used. Each project may have its own unique findings and implications for crop yield prediction using machine learning.

10 FUTURE SCOPE

The future scope of crop yield prediction using machine learning is promising, with several potential areas of development and advancement. Here are some key future directions:

1. Integration of advanced machine learning techniques: As machine learning algorithms continue to evolve, there is potential for incorporating more advanced techniques such as deep learning, neural networks, and ensemble methods. These approaches can potentially improve the accuracy and robustness of crop yield prediction models.

2. Fusion of multi-source data: Integrating diverse data sources such as satellite imagery, remote sensing data, drone-based imaging, and IoT sensor data can enhance the predictive capabilities of crop yield models. By combining data from multiple sources, the models can capture a more comprehensive understanding of the factors influencing crop growth and yield.

3. Real-time and dynamic predictions: The future of crop yield prediction lies in real-time and dynamic forecasting. By continuously updating models with live data streams, weather forecasts, and sensor data, it becomes possible to provide up-to-date and accurate predictions throughout the growing season. This enables farmers to make timely decisions and adapt their management practices accordingly.

4. Improved data collection and quality: Enhancements in data collection methods and technologies can contribute to more reliable and accurate crop yield predictions. Investments in high-quality weather stations, soil sensors, and remote sensing technologies can provide more precise input variables for the models, resulting in improved predictions.

5. Regional and crop-specific models: Developing crop yield prediction models that are tailored to specific regions and crops can yield better accuracy and applicability. Different regions have unique climatic conditions and agricultural practices, and models customized to these specific contexts can provide more relevant insights for local farmers and policymakers.

6. Incorporation of socio-economic factors: In addition to weather and soil data, integrating socio-economic factors such as market prices, supply and demand dynamics, and policy changes can enhance the usefulness of crop yield predictions. This holistic approach can assist farmers and stakeholders in making informed decisions regarding crop choices, marketing strategies, and resource allocation.

7. Integration with precision agriculture technologies: Linking crop yield prediction models with precision agriculture technologies, such as variable rate technology and autonomous farming systems, can optimize resource allocation and improve overall farm efficiency. By leveraging real-time predictions, these technologies can guide on-the-go adjustments and optimize inputs, resulting in improved crop yields.

These are just a few potential future directions for crop yield prediction using machine learning. The field is evolving rapidly, and with advancements in technology, data availability, and algorithmic approaches, there are significant opportunities to enhance the accuracy and practical applications of crop yield prediction models.

11 BIBLIOGRAPHY

[1] Nigam, A., Garg, S., Agrawal, A., & Agrawal, P. (2019, November). Crop yield prediction using machine learning algorithms. In *2019 Fifth International Conference on Image Information Processing (ICIIP)* (pp. 125-130). IEEE.
[2] Medar, R., Rajpurohit, V. S., & Shweta, S. (2019, March). Crop yield prediction using machine learning techniques. In *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)* (pp. 1-5). IEEE.
[3] Champaneri, M., Chachpara, D., Chandvidkar, C., & Rathod, M. (2016). Crop yield prediction using machine learning. *Technology*, *9*, 38.
[4] Crane-Droesch, A. (2018). Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environmental Research Letters*, *13*(11), 114003.

[5] Van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, *177*, 105709.
[6] Nishant, P. S., Venkat, P. S., Avinash, B. L., & Jabber, B. (2020, June). Crop yield prediction based on Indian agriculture using machine learning. In *2020 International Conference for Emerging Technology (INCET)* (pp. 1-4). IEEE.

APPENDIX

App.py:

```python
from flask import Flask, request, render_template
import sklearn
import pickle
import pandas as pd
import os
app = Flask(__name__)    # Initializing flask
# Loading our model:
model = pickle.load(open("RFmodel.pkl","rb"))

@app.route("/")
def home():
    return render_template("home.html")
```

```
@app.route("/predict", methods = ["GET", "POST"])
def predict():
    if request.method == "POST":
        # Nitrogen
        nitrogen = float(request.form["nitrogen"])

        # Phosphorus
        phosphorus = float(request.form["phosphorus"])

        # Potassium
        potassium = float(request.form["potassium"])

        # Temperature
        temperature = float(request.form["temperature"])

        # Humidity Level
        humidity = float(request.form["humidity"])

        # PH level
        phLevel = float(request.form["ph-level"])

        # Rainfall
        rainfall = float(request.form["rainfall"])

        # Making predictions from the values:
        predictions = model.predict([[nitrogen, phosphorus, potassium,
temperature, humidity, phLevel, rainfall]])

        output = predictions[0]
        finalOutput = output.capitalize()

        if (output == "rice" or output == "blackgram" or output ==
"pomegranate" or output == "papaya"
            or output == "cotton" or output == "orange" or output ==
"coffee" or output == "chickpea"
            or output == "mothbeans" or output == "pigeonpeas" or output
== "jute" or output == "mungbeans"
            or output == "lentil" or output == "maize" or output ==
"apple"):
```

```
            cropStatement = finalOutput + " should be harvested. It's a
Kharif crop, so it must be sown at the beginning of the rainy season e.g
between April and May."


        elif (output == "muskmelon" or output == "kidneybeans" or output
== "coconut" or output == "grapes" or output == "banana"):
            cropStatement = finalOutput + " should be harvested. It's a
Rabi crop, so it must be sown at the end of monsoon and beginning of
winter season e.g between September and October."

        elif (output == "watermelon"):
            cropStatement = finalOutput + " should be harvested. It's a
Zaid Crop, so it must be sown between the Kharif and rabi season i.e
between March and June."

        elif (output == "mango"):
            cropStatement = finalOutput + " should be harvested. It's a
cash crop and also perennial. So you can grow it anytime."



    return render_template('CropResult.html',
prediction_text=cropStatement)


if __name__ == '__main__':
    app.run(debug=True)
```