1. Project Name

Yield-Max, an innovative artificial intelligence application, stands as a beacon of sustainable agriculture, striving to address the formidable challenge of achieving Sustainable Development Goal 2 - Zero Hunger.

2. Overview

In the quest for sustainable development, Yield-Max has emerged as a game-changer for agriculture, directly contributing to Sustainable Development Goal 2 - Zero Hunger. Yield-Max harnesses AI, past agricultural data, weather patterns, and soil conditions to provide farmers with personalized recommendations, optimizing crop yields and promoting sustainable practices. By analyzing historical data alongside weather and soil information, the application's recommendation engine guides farmers in choosing the most suitable crops for their specific conditions. This not only maximizes yields but also minimizes resource use and environmental impact, fostering a more resilient and efficient agricultural sector. This AI-powered precision farming tool signifies a significant stride towards a hunger-free world. As we celebrate its first year, the impact on individual farms and the broader goal of global food security becomes increasingly apparent, marking a pivotal moment in the journey towards Zero Hunger.

3. Background

Predicting crop yields is a crucial task for decision-makers at the national and regional levels (such as the EU level) in order to make decisions quickly. Farmers may make decisions about what to grow and when to grow it with the aid of an accurate crop production prediction model.[1] Various methodologies exist for predicting crop yields. Crop yield prediction employs a diverse array of algorithms. The most commonly utilized algorithms include Neural Networks, Decision Trees, Support Vector Machines, Random Forests, Linear Regression, Clustering, K-nearest Neighbor, and ensemble methods such as Forward stagewise algorithm and Gradient boosting tree. Additionally, researchers have explored the effectiveness of multiple linear regression, Extreme Learning Machine, Naïve Bayes, Polynomial regression, Adaptive Neuro-Fuzzy Inference System (ANFIS), and various combinations of these techniques. This diversity reflects the complexity of agricultural data and the need for adaptable models to capture the nuances of different crops and environmental factors. The selection of a specific algorithm often hinges on the nature of the data and the targeted objectives of crop yield prediction models. In terms of features, soil type, rainfall, and temperature are the most commonly used. The variable that is reliant is crop yield. The features were aggregated to provide a more comprehensive perspective of the independent variables (features). Field management, humidity, nutrients, and crop and soil information are the categories into which the independent aspects can be divided. Various evaluation parameters were used as well. The most frequently employed evaluation metrics are Root Mean Square Error (RMSE), R-squared (R2), and Mean Absolute Error (MAE). In terms of validation approaches, cross-validation is predominant, with 10-fold cross-validation being the most frequently employed method. The diversity in evaluation metrics and validation approaches reflects a comprehensive and nuanced assessment of the effectiveness of crop yield prediction models.[1]

4. Key Objectives / Business Objectives

The key objectives of the artificial intelligence application center around addressing critical challenges in agriculture and the contribution to Sustainable Development Goal 2 (Zero Hunger). The foremost aim is to empower farmers by providing tailored recommendations based on historical data, current weather patterns, and soil conditions, ultimately maximizing crop yields. The application is intended to enhance global food security by assisting farmers in choosing the most suitable crops for their specific conditions, fostering sustainable farming practices, and optimizing the use of resources. With a focus on weather-resilient crop planning, the application is designed to enable farmers to adapt to variable weather conditions, mitigating risks and uncertainties. Accessibility for farmers of varying technological backgrounds is ensured through a user-friendly interface, while continuous improvement through data analysis guarantees the relevance and accuracy of recommendations over time. Collaboration with agricultural stakeholders, education and training initiatives, and scalability considerations underscore the holistic approach to achieving measurable impact in promoting sustainable agriculture and alleviating hunger on a global scale. Regular impact assessments and the establishment of key performance indicators ensure that the application's ongoing effectiveness aligns with broader development goals.

a. Research questions

1. What are the key factors influencing crop yields, considering past data, weather patterns, and soil conditions?
2. How can artificial intelligence be effectively employed to provide personalized recommendations for crop selection and maximize agricultural productivity?
3. What are the most relevant evaluation metrics for assessing the accuracy and effectiveness of the AI application in predicting and optimizing crop yields?
4. How does the implementation of the AI application impact farmers' decision-making processes and contribute to achieving Sustainable Development Goal 2 (Zero Hunger)?

b. Key steps

The key steps in developing and implementing the artificial intelligence application for optimizing crop yields involve a systematic approach to data-driven decision-making in agriculture. Commencing with comprehensive data collection encompassing historical agricultural records, weather patterns, and soil conditions, the process advances to the judicious selection and development of machine learning algorithms tailored for crop yield predictions. Feature engineering ensures the extraction of pertinent factors influencing crop productivity, followed by rigorous model training on historical datasets to foster pattern recognition. The definition and implementation of robust evaluation metrics, including Root Mean Square Error (RMSE) and R-squared, serve to validate the accuracy and reliability of the developed AI model. Concurrently, user interface design focuses on creating an accessible and intuitive platform for farmers. Post-implementation, continuous testing, feedback collection, and scalability considerations are paramount, ensuring real-world applicability and adaptability across diverse agricultural landscapes. Collaboration with agricultural stakeholders, outreach initiatives, and the establishment of a monitoring system with key performance indicators contribute to the broader objectives of sustainable agriculture and the application's impact assessment over time, particularly in contributing to the realization of Sustainable Development Goal 2 (Zero Hunger).

5. Methods and Workflow

a. Datasets: SoilHealthDB stands as a pivotal dataset in this project, consolidating soil health measurements gathered from 281 published studies across 41 countries globally. Comprising 5,241 data entries, this comprehensive database encompasses 42 soil health indicators and 45 background indicators detailing factors like climate, elevation, and soil type.

b. Data Cleaning/Preprocessing: Thorough data cleaning and preprocessing were undertaken in the provided code to address missing values, outliers, and inconsistencies, ensuring the dataset's quality and reliability for subsequent machine-learning tasks. The data from the 'SoilData.csv' file was loaded into a pandas DataFrame, initiating an initial exploration of the dataset, including the display of DataFrame contents and the determination of the total number of columns. A subset of essential columns was specified, and the DataFrame has filtered accordingly, with missing columns identified to ensure the inclusion of all relevant features. The sum of NaN values for each column was calculated and displayed using a custom function, 'nan_sum_per_column,' shedding light on data completeness. Strategic approaches were employed to handle missing values, involving the removal of rows with NaN values in critical columns such as 'SiteInfor' and 'SoilFamily' to retain vital data for subsequent analyses. For numerical columns like 'MAT,' 'MAP,' 'Tannual,' 'Elevation,' and 'SoilpH,' missing values were imputed using either the mean of the entire column or group-specific means, contributing to data integrity. Data visualization techniques, such as histograms, and boxplots, were utilized to identify outliers and comprehend the correlation structure among numerical features. Additionally, categorical columns were label-encoded to transform them into a format suitable for machine learning models. The code also established and evaluated machine learning models, including decision trees, random forests, and XGBoost classifiers, trained on the encoded dataset. Class imbalance was addressed through oversampling techniques.

c. Modeling: Random Forest and Decision Trees machine learning algorithms were selected for their suitability in handling complex relationships within agricultural data. The models were trained on the preprocessed datasets, allowing them to learn patterns and relationships for accurate crop yield predictions. Evaluation metrics such as Root Mean Square Error (RMSE) and R-squared were utilized to assess model performance.

d. Deliverables: A user-friendly web interface was created using Flask, HTML, CSS, and Bootstrap to deploy the AI application and make it accessible to farmers over the internet. The web interface will facilitate the interaction with the Random Forest and Decision Trees models, providing farmers with personalized crop recommendations based on historical data, weather patterns, and soil conditions. The application will be regularly updated to incorporate feedback and improvements, ensuring its scalability and adaptability to different agricultural landscapes and crops. The ultimate deliverable is a comprehensive and sustainable solution that empowers farmers to maximize crop yields and contributes to achieving Sustainable Development Goal 2 (Zero Hunger).

REFERENCES

1- 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.* doi:10.1016/j.compag.2020.105709