CROP PREDICTION USING MACHINE LEARNING

Milestone 1: Project Initialization and Planning Phase

The "Project Initialization and Planning Phase" marks the project's outset, defining goals, scope, and stakeholders. This crucial phase establishes project parameters, identifies key team members, allocates resources, and outlines a realistic timeline. It also involves risk assessment and mitigation planning. Successful initiation sets the foundation for a well-organized and efficiently executed machine learning project, ensuring clarity, alignment, and proactive measures for potential challenges.

Activity 1: Define Problem Statement

Problem Statement: Agriculture is one of the main sources of income in India. There is need to improve the sustainability of agriculture with the rate of increase in suicides of farmer due to crop failure and less yield and losses. Hence, it is a significant contribution towards the economic and agricultural welfare of the countries across the world. The Problem Statement revolves around prediction of yield of crops considering different climatic conditions of India including various attributes. Goal of this project is to help the farmers to choose the suitable crop to grow in order to get the required yield and the profit. Need for the crop yield prediction is very much essential at this point of time for selecting the right crop.

Ref. template: Click Here

Activity 2: Project Proposal (Proposed Solution)

Agriculture is a vital sector in many economies, and optimizing crop production is crucial for food security and economic stability. The advent of data science and machine learning has opened up new possibilities for predicting crop yields, enabling farmers to make informed decisions to enhance productivity and sustainability. This proposal outlines a project to develop a crop prediction system using advanced data analytics and machine learning techniques.

Ref. template: Click Here

Activity 3: Initial Project Planning

Primary Goals: Develop a predictive model for crop yields. Enhance decision-making for farmers through data-driven insights.

Secondary Goals: Improve resource allocation (water, fertilizers, etc.).Increase crop yield efficiency and sustainability.

Ref. template: Click Here

Milestone 2: Data Collection and Preprocessing Phase

The Data Collection and Preprocessing Phase involves executing a plan to gather crop prediction data from Kaggle, ensuring data quality through verification and addressing missing values. Preprocessing tasks include cleaning, encoding, and organizing the dataset for subsequent exploratory analysis and machine learning model development.

Activity 1: Data Collection Plan, Raw Data Sources Identified, Data Quality Report

1. **Data Collection:**

1.1 Types of Data:

Historical Crop Yield Data: Data on past crop yields, collected over several years. Weather Data: Historical and real-time data on temperature, precipitation, humidity, wind speed, etc. Soil Data: Information on soil types, pH levels, nutrient content, moisture levels, etc.

1.2 Data Sources:

Government Agencies: Agricultural departments, meteorological departments, soil survey agencies.
Research Institutions: Universities and agricultural research centers.
Commercial Data Providers: Companies providing satellite imagery and weather data.

1.3 Data Collection Steps:

Identify Data Sources: Compile a list of reliable sources for each data type.
Data Acquisition: Obtain datasets through APIs, data downloads, or partnerships.
Data Storage: Use databases (SQL, NoSQL) or cloud storage solutions to store the collected data.
Data Integration: Combine datasets from different sources to create a unified dataset.

2. Data Preprocessing

2.1 Data Cleaning:

Remove Duplicates: Ensure there are no duplicate records in the dataset.
Handle Missing Values: Impute missing values using methods like mean, median, mode, or advanced techniques like KNN imputation.
Correct Inconsistencies: Standardize units of measurement and correct any inconsistencies in data entries.

Ref. template: Click Here

Activity 2: Data Quality Report

The quality of data is critical for the success of any predictive modeling project. This report assesses the quality of the data collected for crop prediction, identifies issues, and recommends steps for improvement.

2. Data Sources

2.1 Historical Crop Yield Data Source: Government agricultural databases

Coverage: Last 10 years

Attributes: Crop type, yield, region, year

2.2 Weather Data

Source: Meteorological agencies, weather stations

Coverage: Daily data for the last 10 years

2.3 Soil Data

Source: Soil survey agencies, research institutions

Coverage: Last 5 years Attributes: Soil type, pH, nutrient levels, moisture content

Ref. template: Click Here

Activity 3: Data Exploration and Preprocessing

2. Data Exploration

2.1 Understanding the Data:

Historical Crop Yield Data: Examine the distribution of crop yields, identify trends over time, and analyze the relationship with geographical regions.

Weather Data: Analyze the variability and patterns in temperature, precipitation, humidity, and wind speed.

2.2 Data Visualization: Visualize the data to identify patterns, trends, and outliers.

2.3 Correlation Analysis: Examine correlations between different features to identify important relationships.

3. Data Preprocessing

3.1 Data Cleaning: Handle missing values, remove duplicates, and correct inconsistencies.

3.2 Data Normalization: Scale numerical features to a common range.

3.3 Handling Outliers: Identify and treat outliers to prevent them from skewing the model.

3.4 Data Splitting: Split the data into training and testing sets for model evaluation.

Ref. template: Click Here

Milestone 3: Model Development Phase

strategic feature selection, evaluating and selecting models (Random Forest, Decision Tree, KNN, SVM), initiating training with code, and rigorously validating and assessing model performance for informed decision-making in the lending The Model Development Phase entails crafting a predictive model for crop prediction. It encompasses process.

Activity 1: Feature Selection Report

Initial Feature List:
Historical Crop Yield Data:
Crop type
Region
Year
Yield (target variable)
Weather Data:
Temperature
Precipitation
Humidity
Wind speed
Ref. template: Click Here

Activity 2: Model Selection Report

The Model Selection Report details the rationale behind choosing Random Forest, Decision Tree, KNN, and SVM models for crop prediction. It considers each model's strengths in handling complex relationships, interpretability, adaptability, and overall predictive performance, ensuring an informed choice aligned with project objectives.

Ref. template: Click Here

Activity 3: Initial Model Training Code, Model Validation and Evaluation Report

The Initial Model Training Code employs selected algorithms on the crop prediction dataset, setting the foundation for predictive modeling. The subsequent Model Validation and Evaluation Report rigorously assesses model performance, employing metrics like accuracy and precision to ensure reliability and effectiveness in predicting crops.

Ref. template: Click Here

Milestone 4: Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase involves refining machine learning models for peak performance. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

Activity 1: Hyperparameter Tuning Documentation

The Random forest model was selected for its superior performance, exhibiting high accuracy during hyperparameter tuning. Its ability to handle complex relationships, minimize overfitting, and optimize predictive accuracy aligns with project objectives, justifying its selection as the final model.

Activity 2: Performance Metrics Comparison Report

The Performance Metrics Comparison Report contrasts the baseline and optimized metrics for various models, specifically highlighting the enhanced performance of the Random forest model. This assessment provides a clear understanding of the refined predictive capabilities achieved through hyperparameter tuning.

Activity 3: Final Model Selection Justification

The Final Model Selection Justification articulates the rationale for choosing Random forest as the ultimate model. Its exceptional accuracy, ability to handle complexity, and successful hyperparameter tuning align with project objectives, ensuring optimal loan approval predictions.

Ref. template: Click Here

Milestone 5: Project Files Submission and Documentation

For project file submission in Github, Kindly click the link and refer to the flow. **Click Here**

For the documentation, Kindly refer to the link. **Click Here**

Milestone 6: Project Demonstration

In the upcoming module called Project Demonstration, individuals will be required to record a video by sharing their screens. They will need to explain their project and demonstrate its execution during the presentation.