**MADRAS INSTITUTE OF TECHNOLOGY**

**DEPARTMENT OF COMPUTER TECHNOLOGY**

**CS6611 CREATIVE AND INNOVATIVE PROJECT**

***CropForEst* - A Machine Learning based**

**crop yield estimation and profitability analysis for Precision Agriculture**

**Second Review**

**Team Members :**

SUNDARAJAN R B - 2020503549

HAMILTON SAMIC S - 2020503511

SURIYAA V - 2020503550

**Project Guide :** Dr. C. Valliyammai

**INTRODUCTION**

The agriculture industry faces a significant challenge in predicting crop yield production and crop demand accurately. The traditional methods of predicting crop yield and demand are time-consuming and can be prone to errors, leading to inefficient use of resources and revenue loss. There is a need for predicting the most suitable crop based on various influential factors and forecasting the profitability.

* To develop an intelligent platform that uses machine learning algorithms to estimate the most suitable crop for cultivation in a given area based on environmental factors.
* To forecast market demand and pricing trends for the identified crop, with the aim of increasing profitability for farmers and stakeholders in the agriculture industry.

**PROJECT MODULES**

This project contains 2 main parts :

1. Crop Prediction
2. Crop Price Estimation

**Data Pre-processing :**

Data is obtained and Pre-processed.

**Model Selection :**

4 Classification models and 3 Regression models were chosen and the most accurate models with least errors were chosen.

**Model Training :**

The chosen models are then trained with various optimization techniques like hyperparameter tuning.

**Model Integration :**

The Model is then integrated into the Application

**Application Interface Development :**

Application Interface (UI) is developed to provide an interactive system for farmers.

**TOOLS :**

**Weather Tracking website :**

AccuWeather is a reliable and comprehensive weather forecasting service that provides users with the most up-to-date weather information. With its advanced forecasting techniques, AccuWeather is a valuable resource for anyone looking to stay informed about the weather.

**OpenStreetMaps API / Google Maps API :**

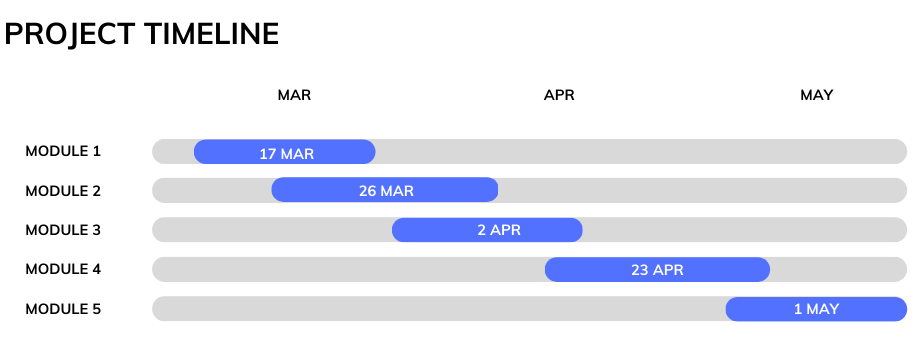
OpenStreetMap (OSM) is a collaborative and free mapping project that provides an open-source and openly licensed database of geographic data. The OpenStreetMap API allows developers to access this data programmatically and use it in their own applications.The OpenStreetMap API provides several different endpoints that allow developers to query the database for different types of geographic data, such as street maps, building footprints, and points of interest.

**Visual Studio Code :**

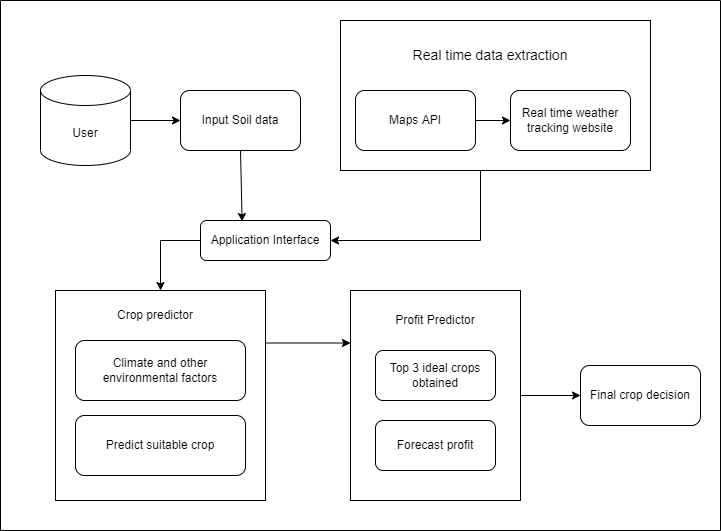
In order to add dynamic and interactive behaviour to the application, scripting in python is required. The scripts are developed in Visual Studio code.

**Google Colab :**

Google Colab is an online platform provided by Google that allows users to write, run, and share code in a Jupyter Notebook environment. The platform offers free access to computing resources such as CPU, GPU, and TPU for machine learning and data analysis tasks. Colab provides a cloud-based environment. Additionally, Colab provides seamless integration with other Google services such as Google Drive. Overall, Colab is a convenient and powerful tool for data science and machine learning tasks.

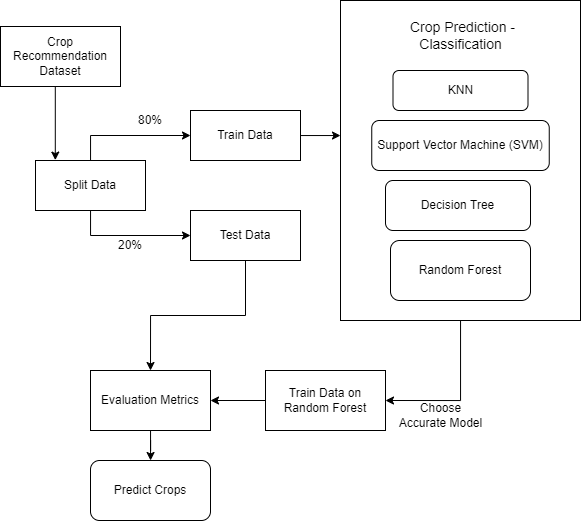


**ARCHITECTURE / TOPOLOGICAL DIAGRAM**

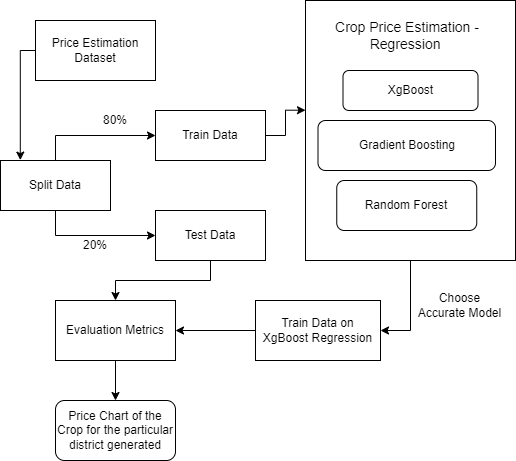


**ALGORITHM FLOW DIAGRAM**

**Crop Prediction**



**Crop Price Estimation**



**IMPLEMENTATION AND RESULTS**

**Data pre-processing**

This module involves data extraction, cleaning, and preparation. Datasets were obtained from Kaggle and new datasets were synthesized using crop data was extracted from https://data.gov.in/ and [www.indiastat.com](http://www.indiastat.com) . The data obtained was cleaned to remove duplicates, irrelevant data, and handle missing values, rounding of values, Removing outliers etc. Normalization or scaling was done to improve the model's convergence and performance. data splitting into training, and testing sets was done.

**Algorithm :**

1. Load the dataset into a Pandas dataframe.
2. Drop irrelevant columns if any.
3. Encode categorical features if any, using LabelEncoder or OneHotEncoder.
4. Convert datetime features to datetime format using pd.to\_datetime().
5. Extract useful features from datetime features, such as month, year, or day of the week.
6. Split the dataset into training and testing sets using train\_test\_split().
7. Scale numerical features if any, using StandardScaler or MinMaxScaler.
8. Apply any other feature engineering techniques such as feature selection or dimensionality reduction.
9. Save the preprocessed data into a new CSV file.

**Model selection and training**

4 Classification models for Crop Prediction and 3 Regression models for Price Estimation were chosen and the dataset was trained with each of these models. The most accurate models with the best fit and least errors were selected finally.

Hyper Parameter Tuning, Feature Engineering were done appropriately.

The Ensemble learning algorithms gave the least errors.

**Classification**

KNN Classifier

Support Vector Machine

Decision Tree Classifier

Random Forest Classifier

**Regression**

Random Forest Regression

XG Boost

Gradient Boosting

**Algorithm :**

1. Load the pre-processed data.
2. Split the data into training and testing sets using a suitable method such as k-fold cross-validation or train-test split.
3. Define the hyperparameters for each candidate model, and the range of values to search over.
4. Perform a grid search or randomized search to find the best hyperparameters for each candidate model, using the training set and a suitable evaluation metric such as mean squared error or R-squared.
5. Evaluate the performance of each model on the testing set using the chosen evaluation metric(s), and select the best performing model as the final model.
6. Optionally, perform feature selection or dimensionality reduction techniques to improve model performance or reduce complexity.
7. Save the final model for deployment in the production environment.

**RESULTS**

**Crop Prediction – Classifiers**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| MODELS | F1 SCORE | ACCURACY | RECALL | PRECISION |
| Random forest classifier | 1.00 | 1.00 | 1.00 | 1.00 |
| Support Vector Machine | 0.98857 | 0.98913 | 0.98502 | 0..99187 |
| Decision tree Classifier | 0.99356 | 0.995 | 0.99230 | 0.99523 |
| KNN Classifier | 0.98040 | 0.98043 | 0.98440 | 0.98269 |

**Crop Price Estimation – Regression models**

|  |  |  |  |
| --- | --- | --- | --- |
| MODELS | MSE | RMSE | MAE |
| Random Forest Regressor | 7902.43 | 88.90 | 70.57 |
| XG Boosting Regressor | 3665.25 | 60.54 | 51.48 |
| Gradient Boosting Regressor | 3704.94 | 60.87 | 51.55 |

**Model Integration :**

The trained models are then exported as .joblib file in colab. This can be integrated into the UI with the help of javascript.

**Application interface development :**

This will involve developing an application interface to provide crop prediction and profitability forecasting functionality. The trained models are integrated into the UI and the farmers are able to see the crop analysis.