

CIS 890 Introduction to Food Informatics

Revanth Babu Raavi and Prof. Dr Hande Kucuk McGinty

EXPLORING CROP YIELD PREDICTION USING MACHINE LEARNING TECHNIQUES

Progress Report

Non-technical Summary: Climatic and Soil characteristics play a major role in agriculture in order to produce more yield. But traditional farming practices often involve planting the same type of crops, using excessive fertilizers, and following conventional wisdom. This project aims to use Machine Learning model to predict the wheat yield so that it will be beneficial for the farmers by meeting their requirements.

Objective: The main objective of this project is to estimate crop yield using varied machine learning techniques which provides an accurate prediction. It involves three steps. The first step is to collect historic data and the second step is applying varied machine learning regression algorithms to that data and final step is deploying an application to predict the wheat yield using the regressor algorithms .

Status of Work:

- As first step is collecting data, The data needs to be provided by domain expert. Unfortunately it didn't happen. To address this issue, an open source data was collected from kaggle. However, the newly acquired data lacked measuring units, prompting the need to make assumptions to compensate for this absence.
- A wide variety of regressor algorithms(models) like Support Vector Regressor(SVR), K-Nearest Neighbor Regressor, Bagging and Extra-Tree Regressor are taken. The reason why regression is used because it involves predicting continuous numerical values and binds relationship between input variable(features) and continuous target variable.
- XAMPP server is used to run the port that connects with the databases.
- SQL Yog, which is a graphical user interface(GUI) tool that manages and interacts with MySQL databases is used.

Approach: It starts with importing essential libraries for data processing, machine learning and data visualization. For this, I used Google Colab to process the regressor algorithms. The collected data is loaded as a CSV file and preprocessed with missing data being handled and splitted the features into training and testing sets. Several regressor algorithms were applied on the data making them to get trained on training data, used to make predictions on the test data, and evaluated the performance of each model using metrics such as R-Square and Time factor. Finally, the code generates visualizations of the model's performance by plotting residuals, enabling an evaluation of how accurately the model's predictions match the actual values. The best performing is saved to predict the yield of the crop. Visual Studio Code editor is used for connecting the front end and backend part and to run the server. For the front end web application part, HTML is used with the help of FLASK at the backend. XAMPP server is used along with SQL Yog Enterprise, which connects to the databases. Front end web application has all the data that needs to be given by the user manually to get the desired Output. Making an extension, I am trying to incorporate this model to work for different crops.

Status of Deliverables: As a part of the process, several models have been compared, evaluated, and the most optimal one has been preserved. Meanwhile the user interface is prepared and the process of GraphDB is in progress.

Status of Outcomes: The project is in final stages of completion and effectively predicting yields. This project will help the farmers and analysts to get an idea about yielding of crops based on climatic and soil characteristics and helps to increase production of the yield while planting in the future.