CIS 890 Introduction to Food Informatics

Revanth Babu Raavi and Prof. Dr Hande Kucuk McGinty

Non-technical Summary: Climatic and Soil charactersitics 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 ML model to predict the yield of the wheat so that it will be beneficial for the farmers by meeting their requirements.

Objective: The main objective of this project 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 best regressor algorithm which is accurate.

Approach: For this to be achieved, we use regression algorithms. To do so, below are the steps to be followed

- Data Preprocessing - In this the missing data needs to checked and handled. This will be done by removing null values , rows/columns with missing data. The data has then to be encoded into numeric values(strings to numbers) if variables are found ,as models are effective with numeric data.
- Splitting the features: The dataset needs to be splitted into two parts . One is for Training and other is for Testing . There are two variables X and Y in each of them, X is the feature and to predict the Y target and same happens for testing also.
- Model selection: Choosing the regression algorithms that has to be applied on independent variables and the target variable.
- Model evaluation: Evaluating the performance of each model on the Testing data. Common evaluation metrics include R square and Time Factor.

Deliverables: The main deliverable of this project is to deploy an application to predict the yield of the wheat with the help of regression algorithms. The code and application needs to be submitted.

Outcomes: The main outcome is to predict the yield of wheat so that it will be useful for farmers and analysts to select the crop and tries to increase more production based on the output that this model gives. By taking the outcome into consideration, further developments can also be made like applying to larger information and adding more features in order to increase the yield.