Marcus McKenzie

Project

DroTek (IBM) Case-Study

This case study will be centered around the idea of optimizing agriculture production and profit for farmers, using Iot devices. This is currently one of IBM's projects, in which the company is working with DroTek to use large amounts of various data to optimize agricultural production and profit, in order to meet an increased future demand of food.

My proposed solution will involve analyzing factors and unstructured data to determine which factors are significant in crop production. These factors and crop data will then be implemented into a regression model. This data, along with the agricultural constraints, will then be provided to an optimization model which will provide a recommendation of the best time for a farmer to harvest their crops to maximize production and profit.

The information below demonstrates the combination of the models, the data that will be necessary for this type of analysis, how often the data and models will have to be refreshed and the possible problems with this solution. Overall, the hope is to develop an analytics model to determine a potential prescriptive solution for improving agriculture yields for farmers.

I. The company's overarching goal is to determine the best time for a farmer tio harvest in order to maximize the yield production and profit for the farmer. This can be broken down into three parts:

    A. **Determine important factors for crop high yield**
    B. **Predict future price and demand of crops**
    C. **Determine best time to harvest**

II. Determine important factors for high crop yields: the company would like to have a better understanding of which factors impact crop yield. This information will help them in predicting future production and could also be used to provide insights to the farmers about the best methods/actions that they can take in the future to increase their crop production.

To predict which factors are the most important in creating high production and yields, crop and environmental data from previous years will have to be analyzed. Some examples of this data are crop size, crop color, temperature, light, air pressure, precipitation, groundwater and crop yield. This and more data is collected and stored using Iot devices, developed by the DroTek team. Using this data, we can then implement a Lasso model to determine which of these factors has the largest impact on crop production. This information can then be used to predict future crop production and also provide information to the farmers about what factors to focus on to improve future crop production.

Some of this data, such as crop size, precipitation and air pressure, will likely have to be collected daily. While other data, such as crop production, can likely be collected each harvest. In order to provide a higher accuracy for the following season, the data should be refreshed and rerun after each harvest season, with the most recent data added to the previous data. This should allow the models to utilize new information and notice any trends that may be occurring.

Step 1:

Given: Previous Crop, Environment, Production and Demand data
Use: Lasso Variable Selection
To: Determine important factors for high crop production/yields

III. Determine future production and crop prices: The company would benefit from providing the farmer's insight into crop production and prices in the future. Obtaining this information will allow the company to determine when the most optimal time to harvest the crops would be, in order to maximize the production and profit for farmers.

To predict the future crop production for the farmers, the company can utilize the factors that were determined to be important from the Lasso model and the crop data as information that is then used in a regression model. This should provide a fairly accurate prediction of the amount of crop production for that season. In order to determine the crop price in the future, the company can implement a similar model, using data about expected demand and crop production for the season. This should provide insights into

the amount of return that the farmers will make and help to determine the best time to harvest the crops.

The data that is used for these models could be collected each harvest season. However, It would be beneficial to be able to have crop production and demand data from previous years to obtain a larger sample size. The season and regions that the data is collected during will also have to be noted as this can likely affect the results, especially if taken from different farms or different regions of the world. In addition, this data would likely have to be refreshed, updated and the model rerun for up to date predictions each season.

Step 2:

Given: important factors, current/previous crop data and crop prices
Use: Linear Regression or Holt-Winters
To: Predict future crop production and future price of crops

IV. Determine best time to harvest crops: using the predictions from the regression model, the company will be able to provide insights on when the best time to harvest the crops.

In order to determine when the best time to harvest is, the company could use an optimization model that optimizes for the maximum amount of production and return for the farmers. For the optimization model, the future production, prices and demand would be important. The model would also have to take into account the agricultural constraints of things like the amount of time to harvest and transport time from farm to factory.

It may be beneficial to refresh and rerun this model multiple times throughout the season, as the predictions and constraints could change depending on current events and weather.

Step 3:

Given: Crop production predictions, crop price predictions, crop data, Agricultural constraints
Use: Optimization
To: Determine best time to harvest crops for high production and return