Crop Yield Prediction

The aim of this notebook is to predict the crop yield using the data from the given dataset. The dataset has been taken from the Crop Yield Prediction.

Conclusion

From the above analysis in the Exploratory Data Analysis, it is clear that there are two crops in the dataset, whose presence has been confirmed with presence of two definite clusters in the graphs of rainfall, temperature, and crop yield. The dataset has been collected for two different crops. The dataset has signs of proportional relationship between the nutrients and the crop yield. However, the relationship between the crop yield and the other columns is not directly proportional. This could be due to the other factors such as soil type, weather conditions, and crop breed. On the whole, the dataset is not much complex to make concrete decisions based on the graphs only. Coming to the machine learning models, I have used Decision Tree Regressor and Random Forest Regressor to predict the crop yield. The Random Forest Regressor has performed better than the Decision Tree Regressor. The Random Forest Regressor has a R2 score of 0.802 and the Decision Tree Regressor has r2 score of 0.77. From the feature importance graph, we can see that the temperature has the highest importance in predicting the crop yield. The temperature is followed by the rainfall, however the macro nutrients tend to have comparatively less importance in predicting the crop yield.