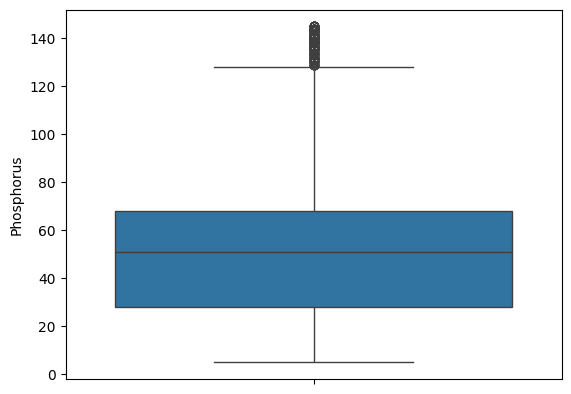
**Machine learning Project in selecting crop according to soil type**

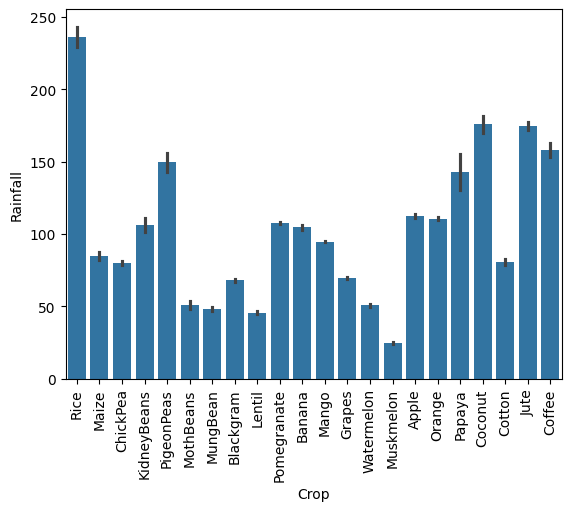
This data (2200 rows) has soil features such as Nitrogen, Phosphorus, Potassium, Temperature, Humidity, pH Value, Rainfall. Here the target column is having name of different crops such as rice, maize, banana, mango etc. there are total 21 such crop names. So, it is multinomial classification problem.

Firstly, I did some data analysis both univariate and bivariate. I have taken the independent columns at a time and found that there are outliers in the data.

This the boxplot of phosphorus



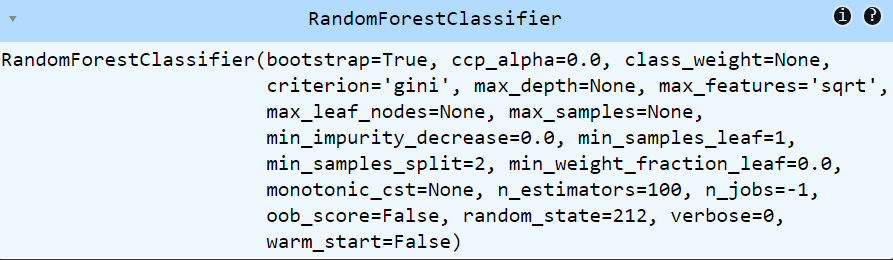
Going forward I have done bivariate analysis taken rainfall and crop name target column and got this graph.

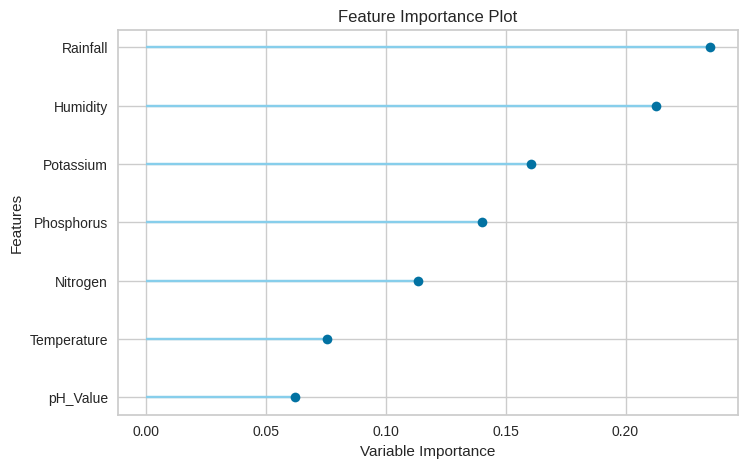


Then I converted the target column into numerical for training the model.

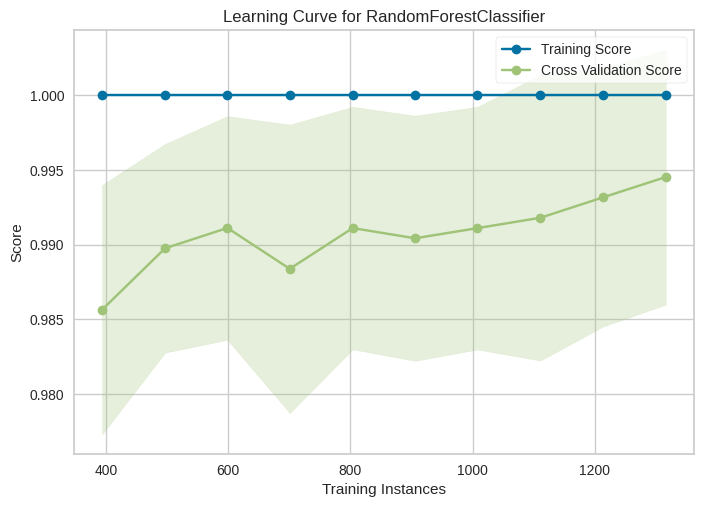
With help of pycaret I have analyzed which algorithm giving good results and found that random forest is performing well in both without preprocessing and with preprocessing. Then I created the model, trained the model and predicted the model.

The parameters which help to make optimized model are this.





To predict the model this features (rainfall, humidity, potassium etc.) are helping more to predict the target column. The evaluation graph such as learning graph



I have used another algorithm that is naïve bayes for prediction purpose. I have chosen this algorithm because it is giving second highest results after random forest. I have done same procedure as I did in before algorithm. I found slight difference in the evaluation diagrams that is learning graph is quiet different from random forest learning graph.

The diagram below

