CROP ANALYSIS AND PREDICTION

By Sruti Lakshmi Narasimhan

About the Dataset

This dataset was build by augmenting datasets of rainfall, climate and fertilizer data available for India.

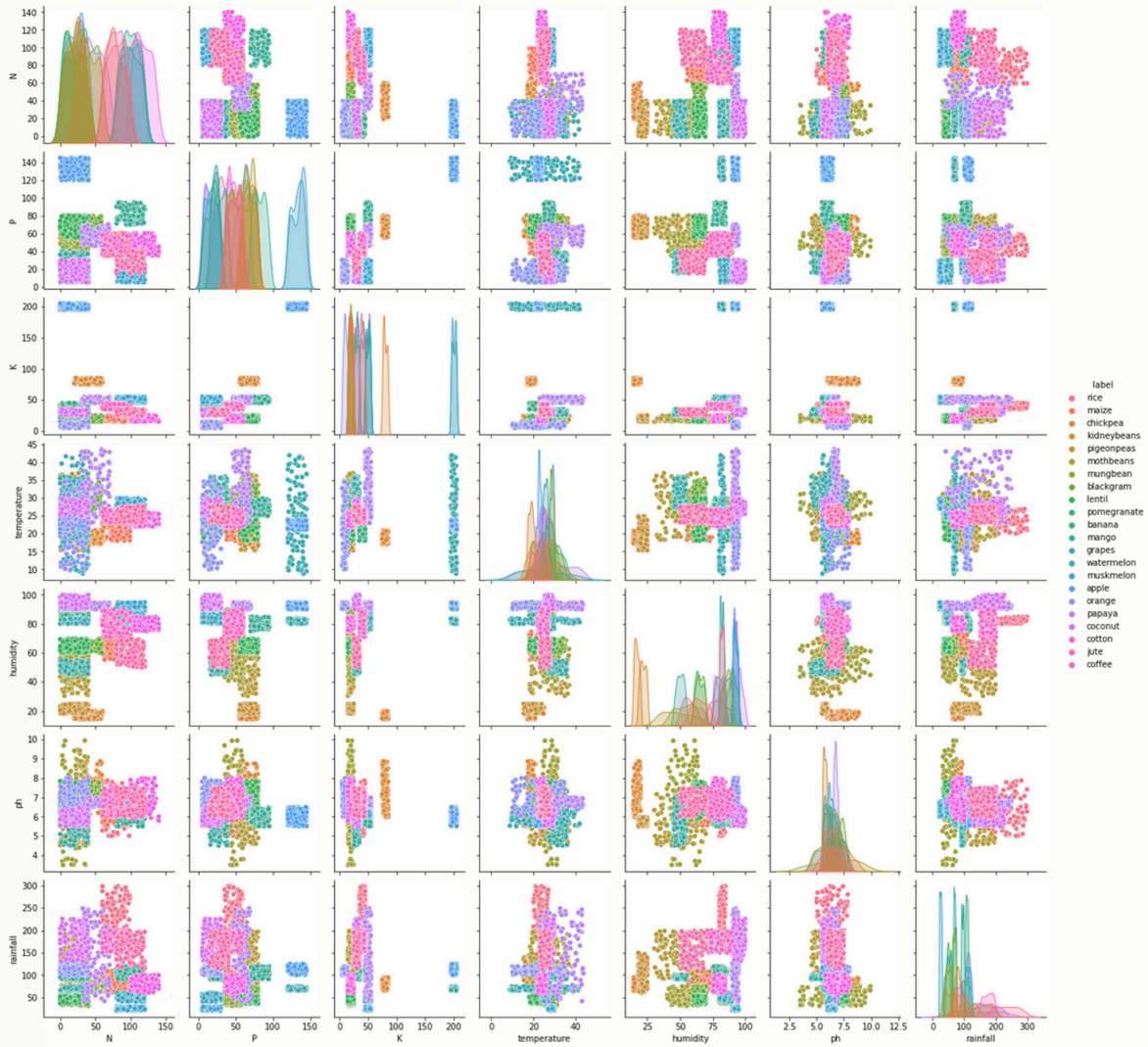
Data fields

N - ratio of Nitrogen content in soil
P - ratio of Phosphorous content in soil
K - ratio of Potassium content in soil
temperature - temperature in degree Celsius
humidity - relative humidity in %
ph - ph value of the soil
rainfall - rainfall in mm
label- suitable crop

Data Visualization

There are 22 different crops and they are rice, maize, chickpea, kidneybeans, pigeonpeas, mothbeans, mungbean, blackgram, lentil, pomegranate, banana, mango, grapes, watermelon, muskmelon, apple, orange, papaya, coconut, cotton, jute, coffee

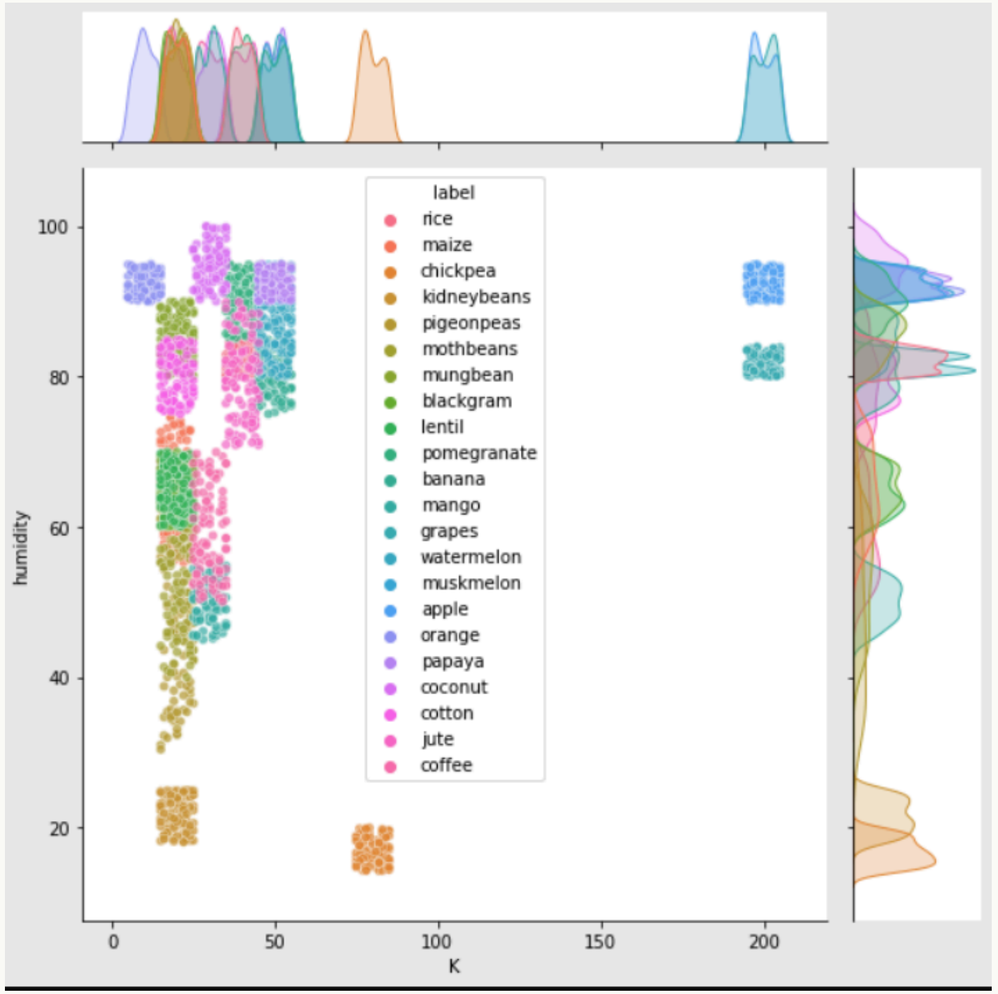
Here is a pair plot of this dataset it is a pairwise relationships in a dataset.



Data Visualization

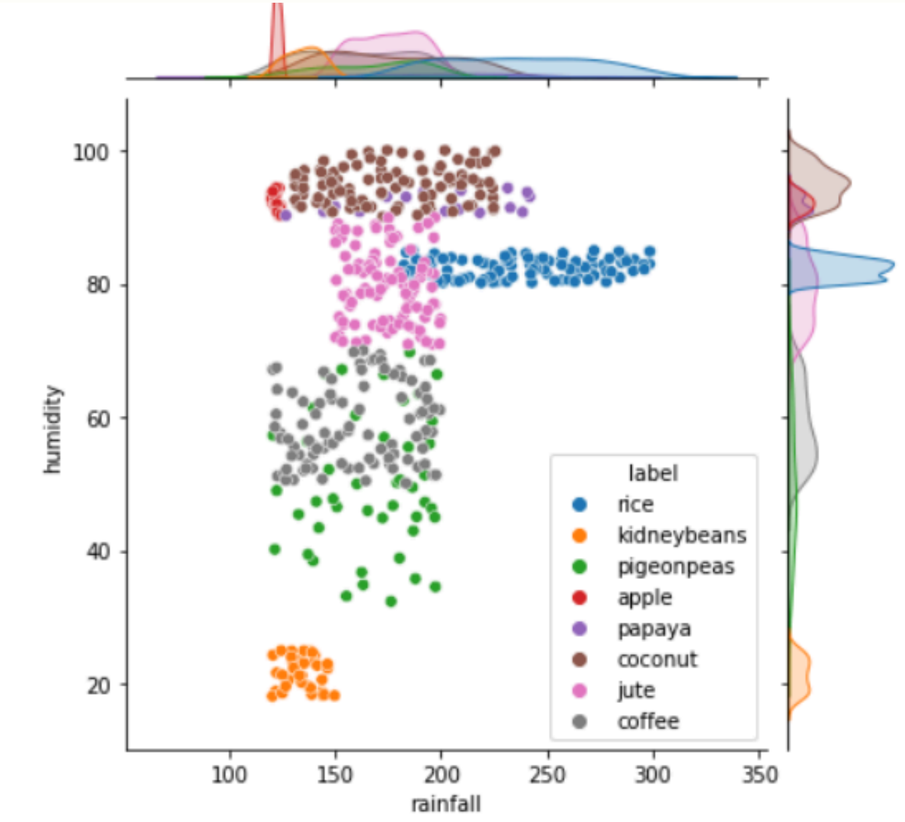
Here is a specfic case of pairplot between `humidity` and `K` (potassium levels in the soil.)

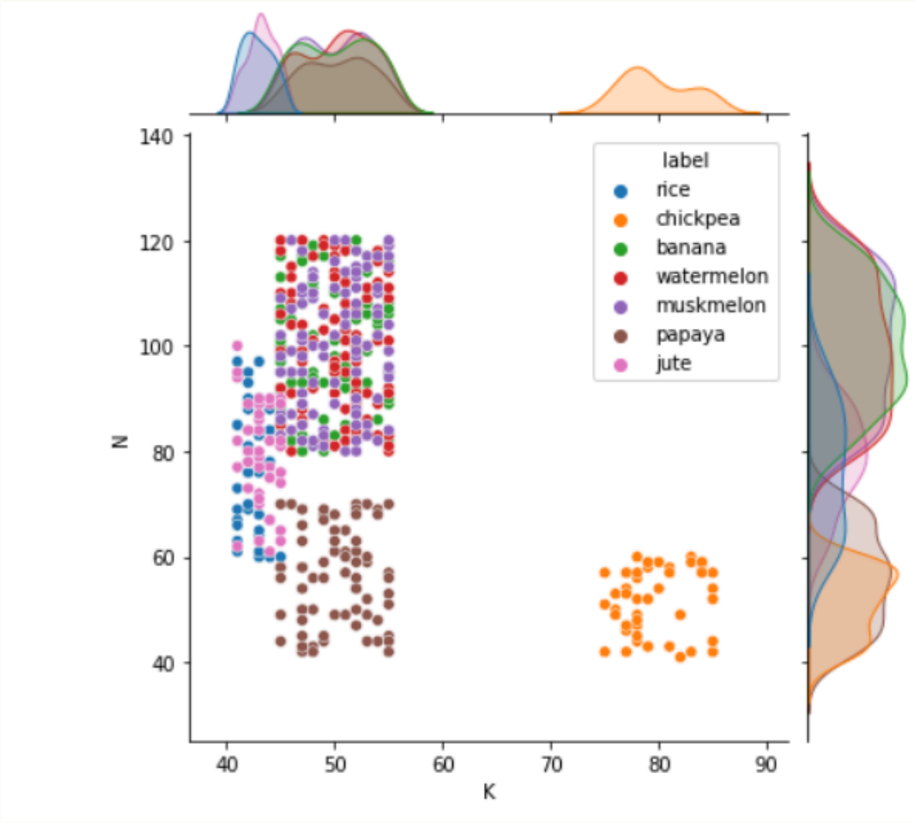
sns.jointplot() can be used for bivariate analysis to plot between humidity and K levels based on Label type. It further generates frequency distribution of classes with respect to features



Data Visualization

The graph in the right depicts the rainfall and humidity required for a crop and the left graph correlates with average potassium (K) and average nitrogen (N) value (both>50).
These soil ingredients direcly affects nutrition value of the food. Fruits which have high nutrients typically has consistent potassium values.

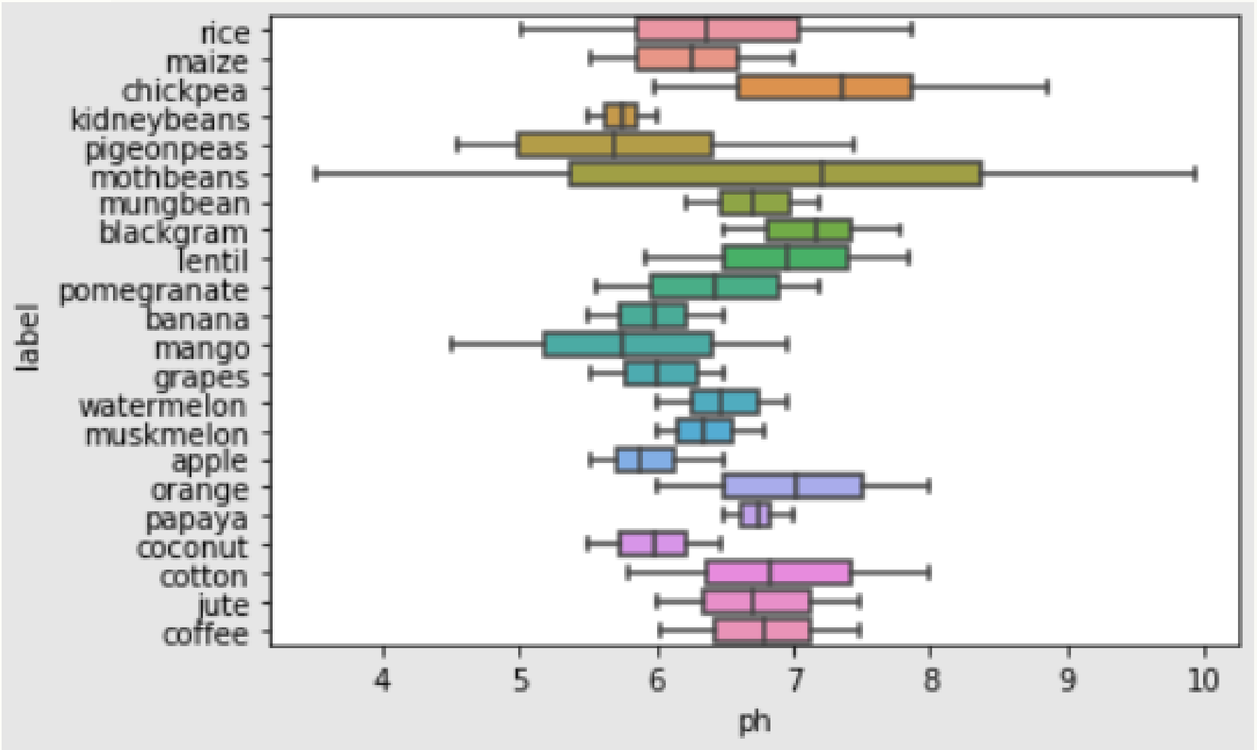




Data Visualization

The ph values are critical when it comes to soil. A stability between 6 and 7 is preffered.

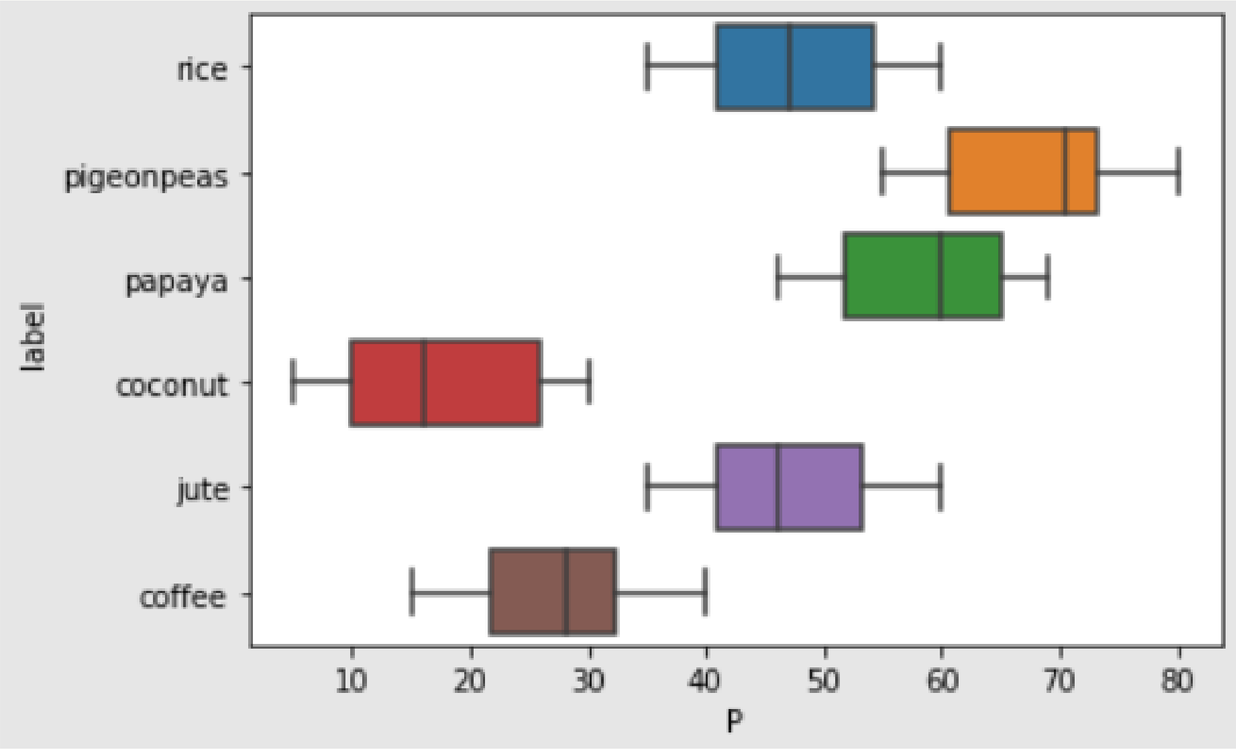
We can see most of the crops have ph between this range.



Data Visualization

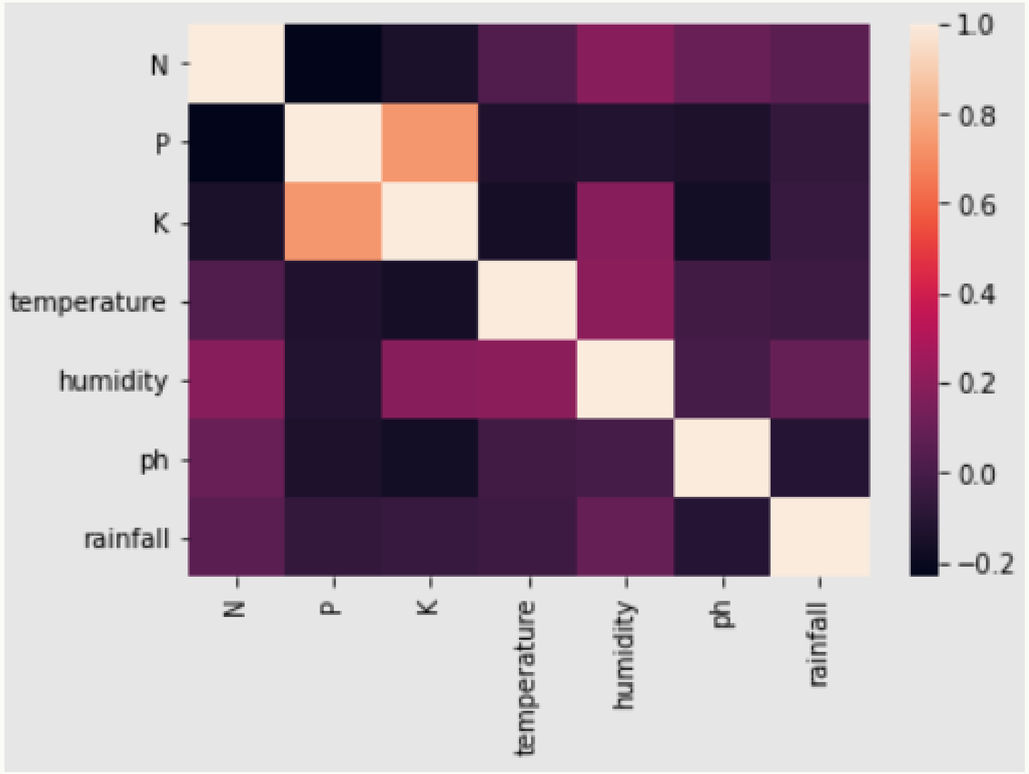
Another interesting analysis where Phosphorous levels are quite differentiable when it rains heavily (above 150 mm).

We can see most of the crops have ph between this range.



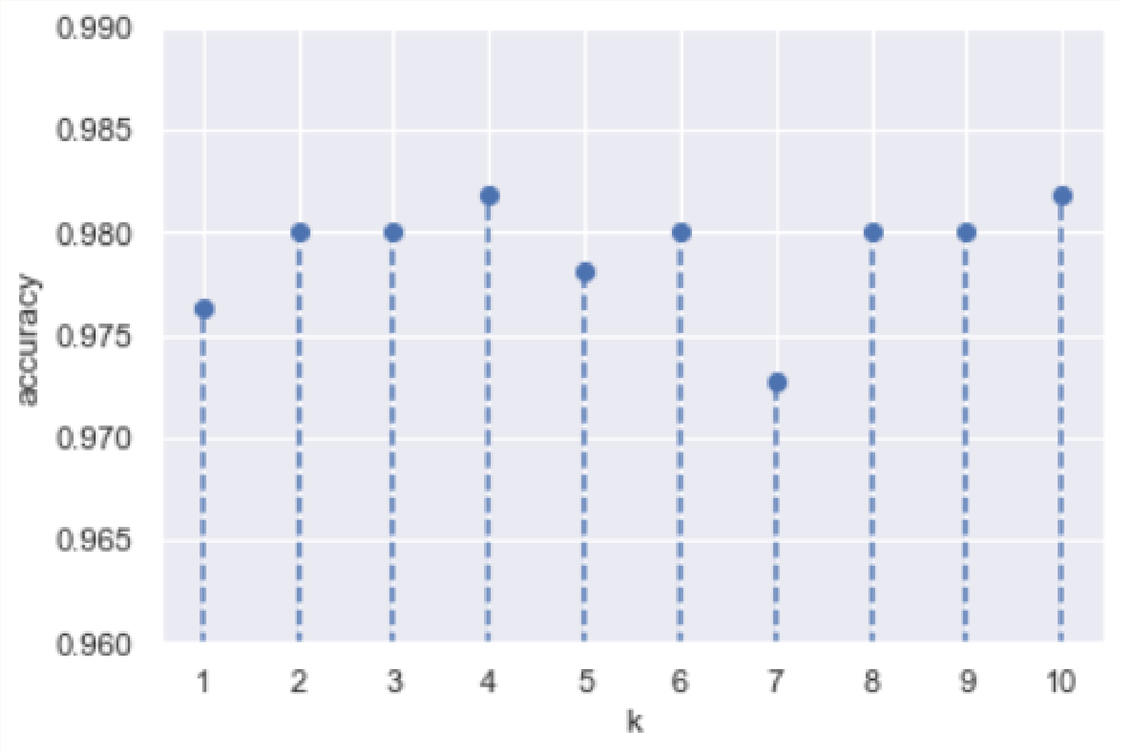
Data Visualization

Correlation visualization between features. We can see how Phosphorous levels and Potassium levels are highly correlated.



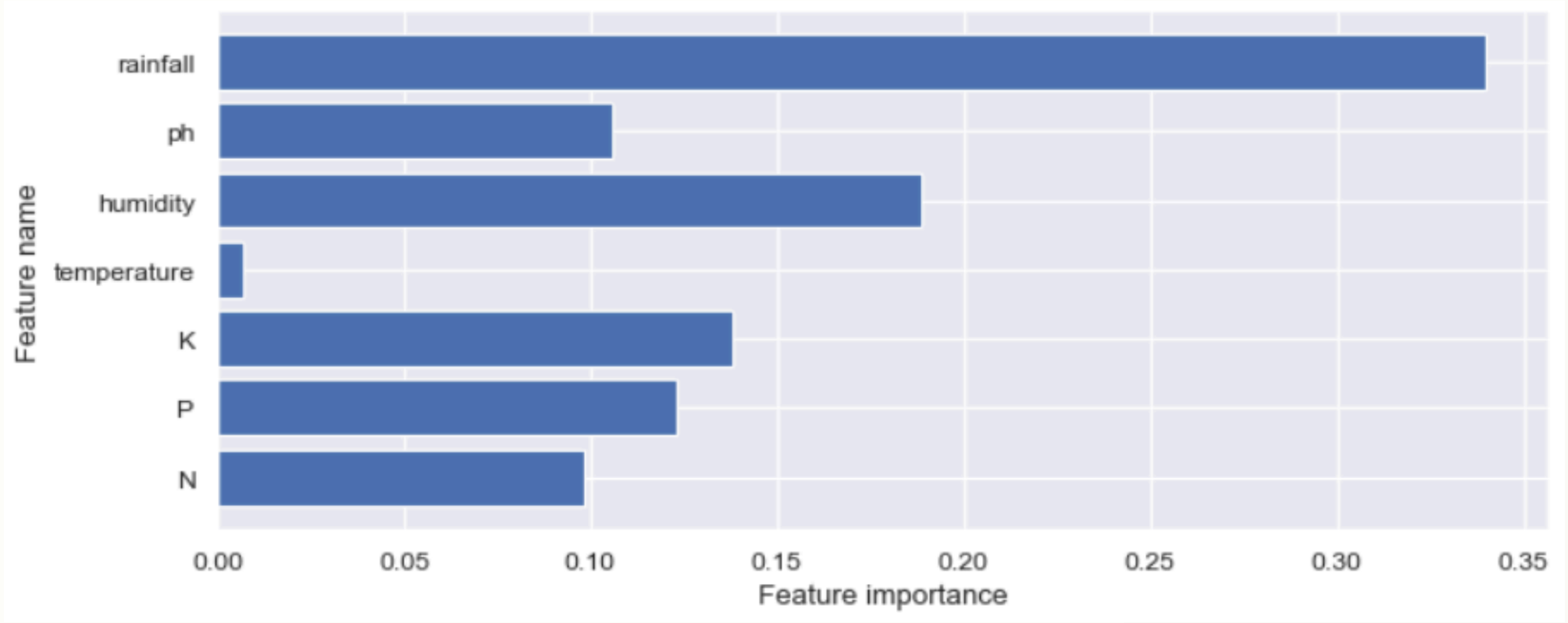
ML model: KNN

To find different values of n_neighbors to fine tune and get better results



ML model:
Decision tree

Here we are visualizing the import features which are taken into consideration by decision trees.



Other ML models used:

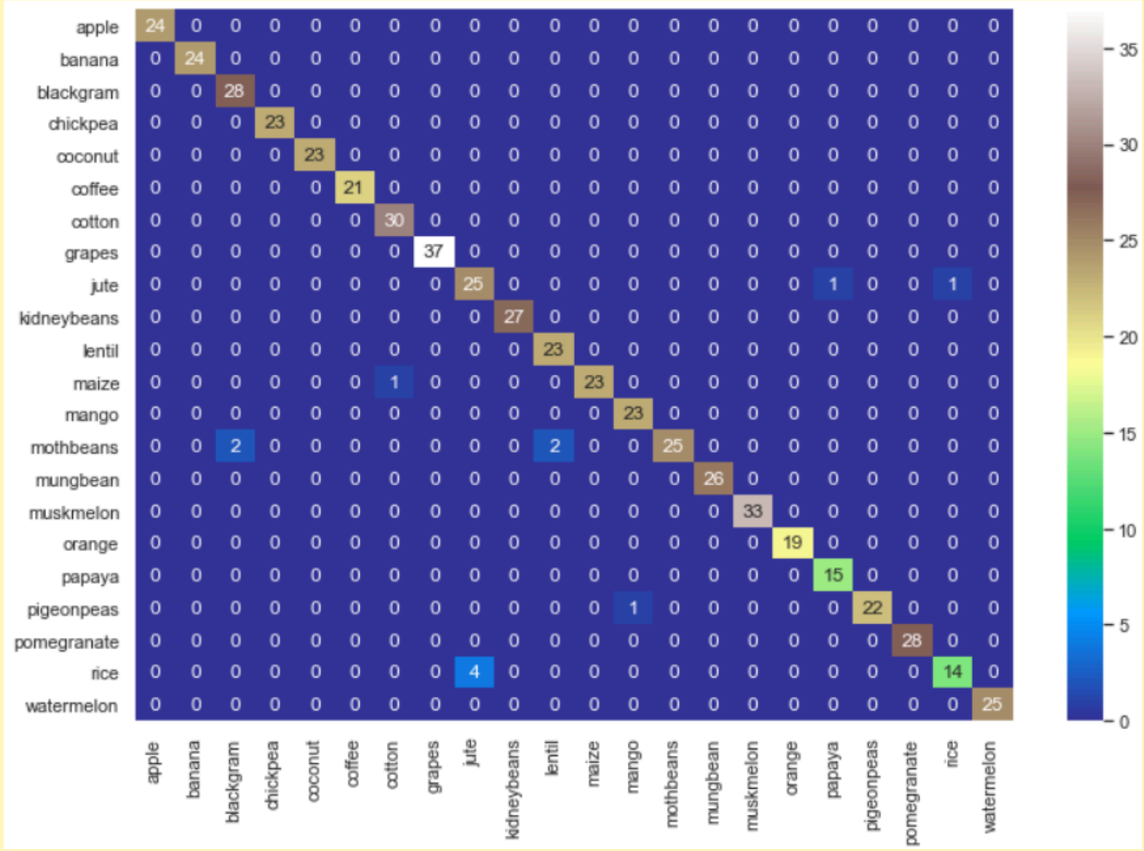
- SVC
  - Linear
  - Rbf
  - Polly
- Random Forest
- Gradient boosting

Accuracy

KNN	SVC (poly kernel)	Decision Tree	Random Forest	Gradient Boosting
0.9781818	0.9890909	0.9872727	0.97	0.9945454

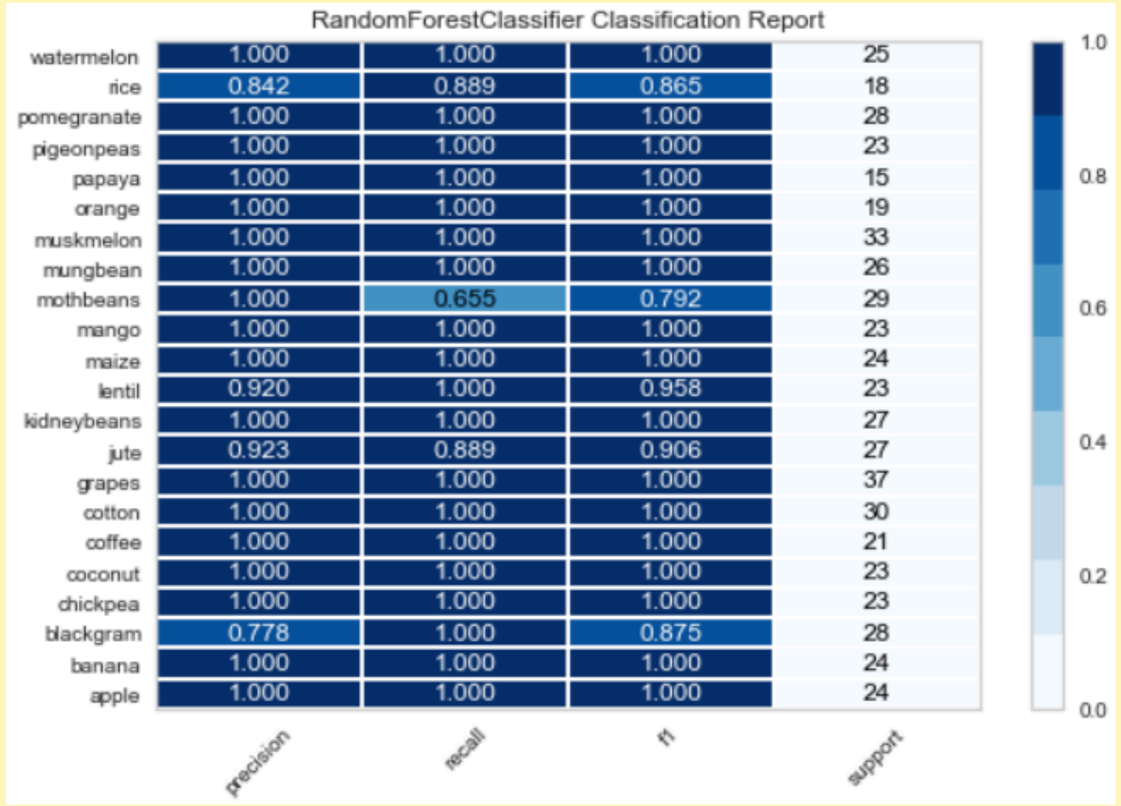
Confusion Matrix

Apply page animations and transitions to emphasize ideas and make them even more memorable. Find the magic and fun in presenting, too, by pressing C for confetti, D for a drumroll, and O for bubbles.



Classification Report

We are using yellowbrick for classification report as they are great for visualizing in a tabular format



THANK YOU