**TOPIC:** ENABLING FARMERS TO MAKE INFORMED DECISIONS BASED ON RELIABLE DATA.

Submitted by**:** MAYANK CHONDE

**ABOUT THE DATA:**

Precision agriculture is rapidly becoming the go-to farming strategy, enabling farmers to make informed decisions based on reliable data. Leveraging a predictive model built on a comprehensive dataset allows farmers to select the best crops to grow on their specific farm, considering various parameters.

**CONTEXT:**

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

DATA FIELDS:

N – The ratio of Nitrogen content in the soil.

P – The percentage of Phosphorous content in the ground.

K – The percentage of Potassium content in the ground.

Temperature – the temperature in degrees Celsius.

Humidity - relative humidity in %.

ph - ph value of the soil.

Rainfall - rainfall in mm.

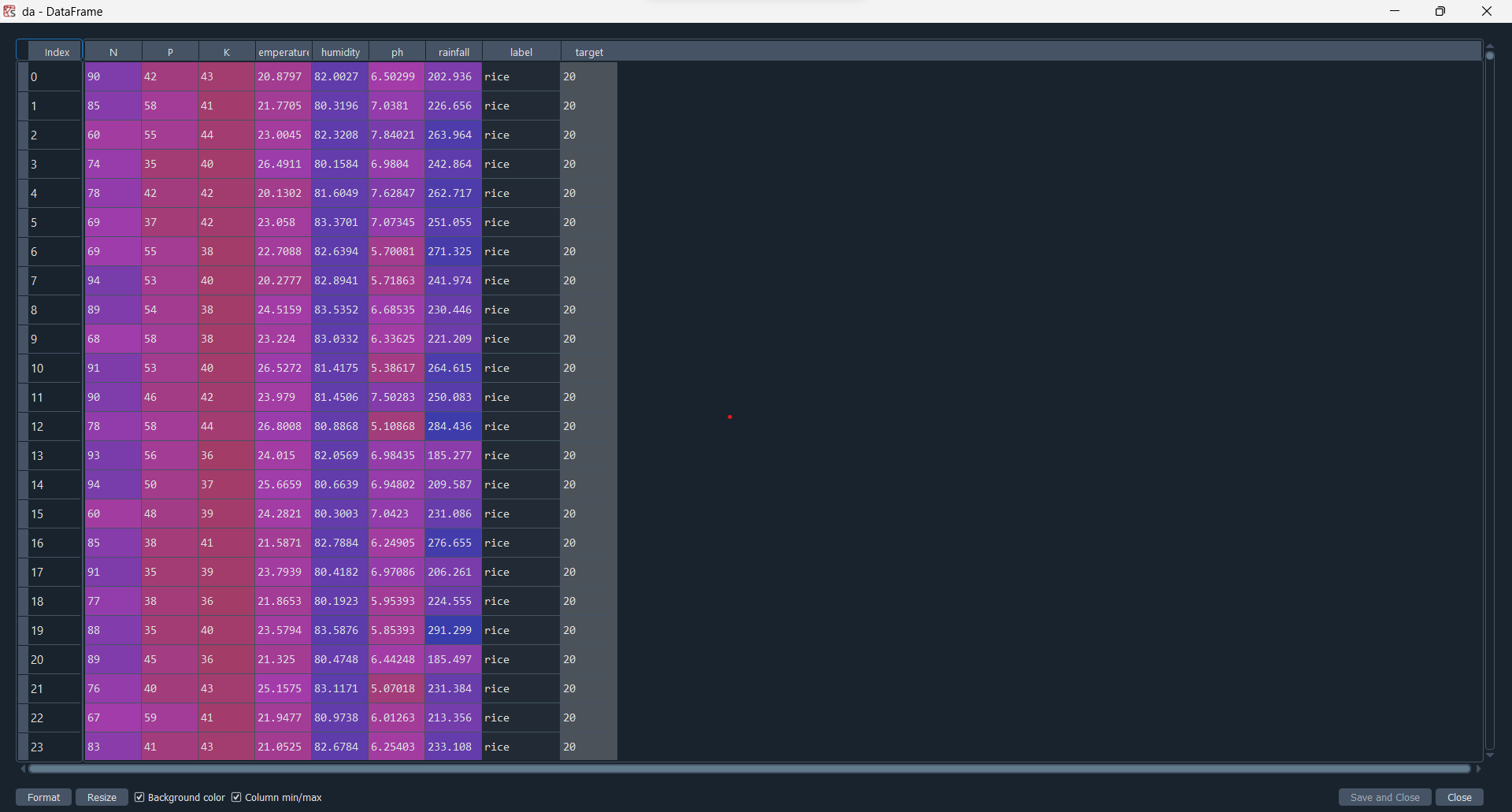
**RESULTS AND DISCUSSION:**

**IMPORTING DATASET:**

**import** pandas **as** pd

da**=**pd**.**read\_csv**(**"D:\Lecture\Data Mining\Lab Exercise 10\Crop\_recommendation.csv"**)**

**OUTPUT:**

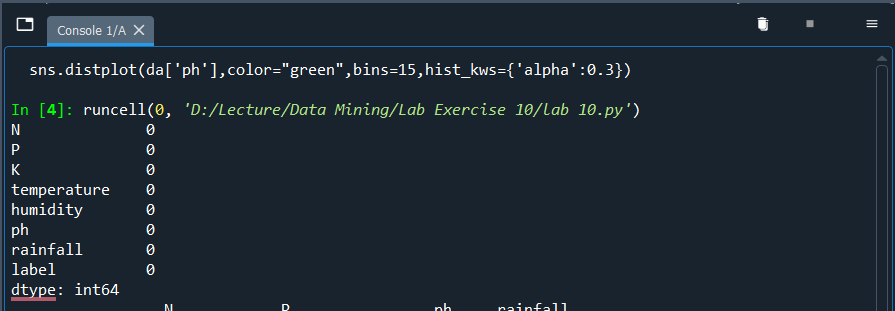


**CHECKING NULL VALUES:**

#checking null values

**print(**da**.**isnull**().**sum**())**

**Output:**

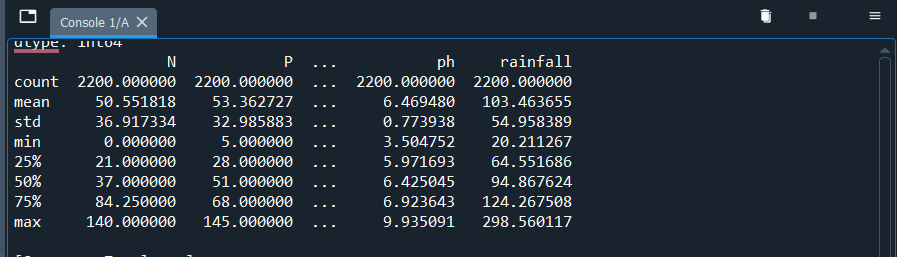


**DESCRIBE THE DATA:**

#decribe the data

**print(**da**.**describe**())**

**Output:**



**PLOTTING DISTRIBUTION OF TEMPERATURE AND PH.:**

#Plotting distribution of temperature and ph.

**import** seaborn **as** sns

**import** matplotlib**.**pyplot **as** plt

plt**.**figure**(**figsize**=(**12**,**5**))**

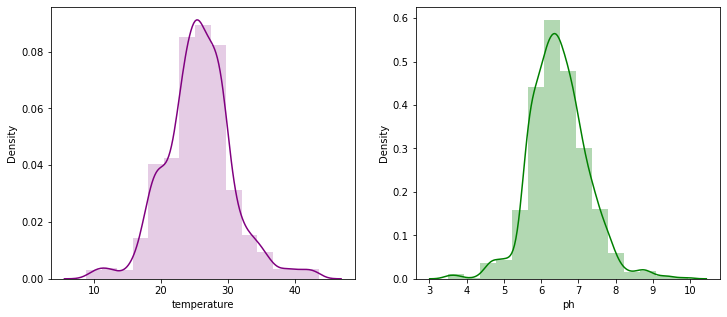
plt**.**subplot**(**1**,** 2**,** 1**)**

sns**.**distplot**(**da**[**'temperature'**],**color**=**"purple"**,**bins**=**15**,**hist\_kws**={**'alpha'**:**0.2**})**

plt**.**subplot**(**1**,** 2**,** 2**)**

sns**.**distplot**(**da**[**'ph'**],**color**=**"green"**,**bins**=**15**,**hist\_kws**={**'alpha'**:**0.3**})**

**OUTPUT:**



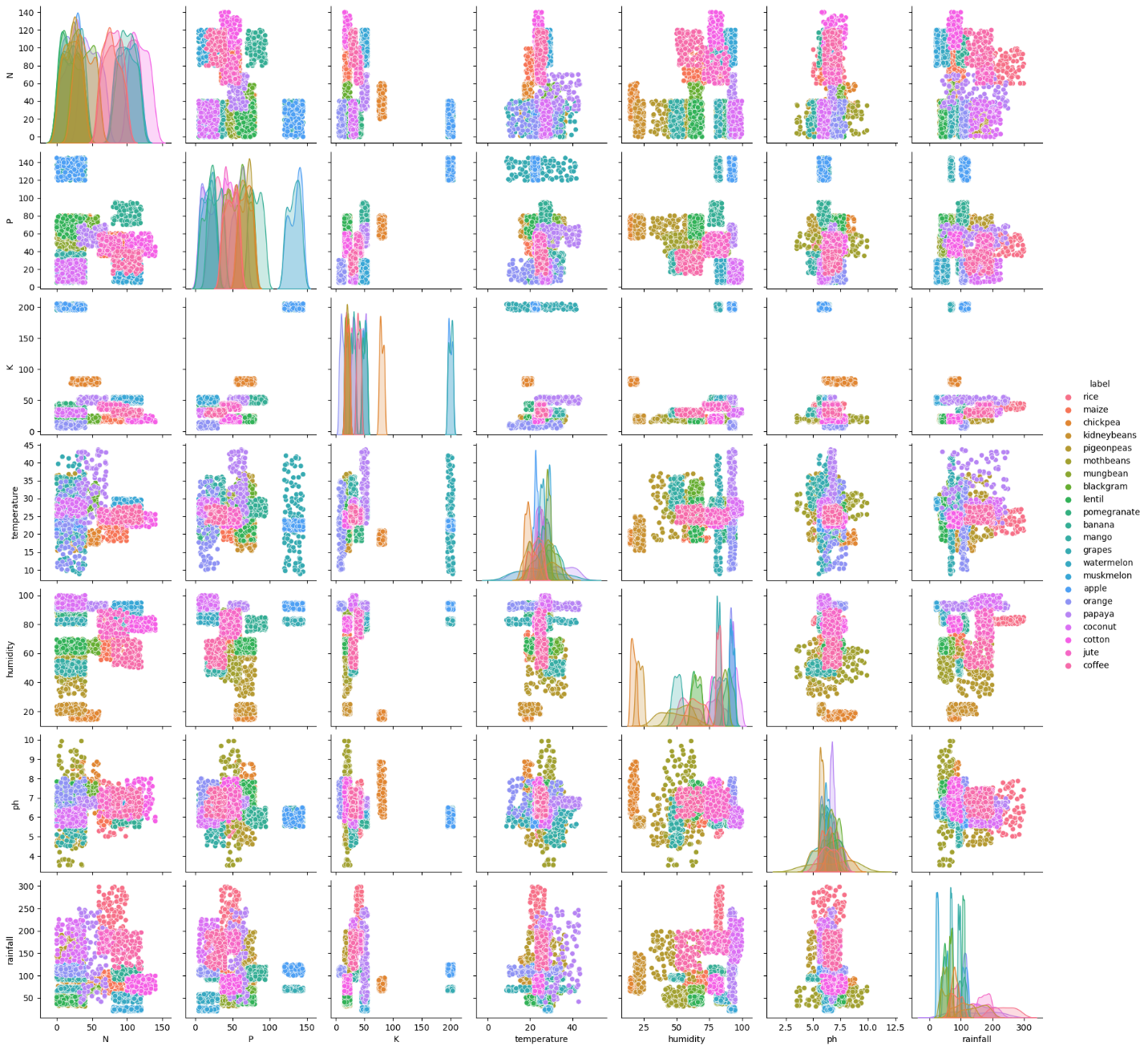
**INTERPRETATION:**

The bell curve describes the range of values that can be expected from a data set. By looking at the shape of the curve, it is possible to understand how likely it is that a data point will be close to the mean or far away from it.

**PAIR PLOT FOR LABEL:**

#plotting pair plot

sns**.**pairplot**(**da**,** hue **=** 'label'**)**

**OUTPUT:**

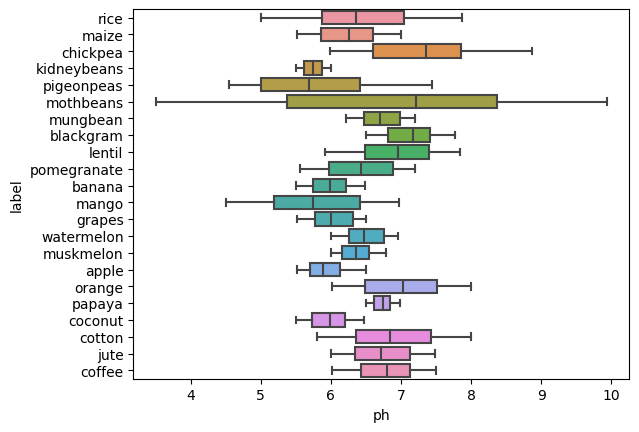
**INTERPRETATION:**The rainy season is a crucial period for crop production in India. During this season, the average rainfall is 120mm, and the temperature is mild, rarely exceeding 30 degrees Celsius. While this might not be the ideal condition for certain crops, it benefits this is why others. Rice and coconut are two crops that are well-suited for the rainy season. Rice is a staple crop in India, requiring rainfall and high humidity for optimal growth. This is Inst the coast is a major rice producer, with an average annual rainfall of 220 mm. As the monsoons bring in the necessary rain, the soil moisture increases pH level also changes, making it ideal for rice growth. In the same way, coconut, with most of the production coming from the coastal areas.

The monsoons bring in ample rainfall and humidity, making these regions ideal for coconut production. The wet season is necessary for crop production in India, and it is essential to ensure the right.

**BOXPLOT:**

sns**.**boxplot**(**y**=**'label'**,**x**=**'ph'**,**data**=**da**)**

**OUTPUT:**



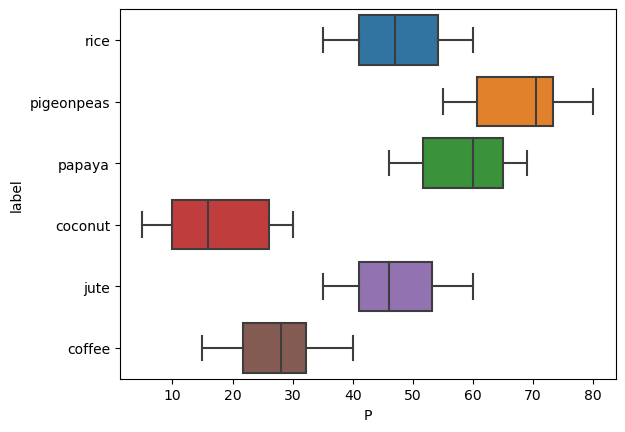
**INTERPRETATION:**

The pH of the soil is essential in determining its fertility and ability to support plant life. Generally, a soil pH between 6 and 7 is ideal for most plants as it provides the right balance of nutrient availability and soil acidity. Too low a pH lead to nutrient deficiencies, while too high a pH can lead to toxic levels of certain elements.

**BOXPLOT WITH RESPECT TO RAINFALL:**

sns**.**boxplot**(**y**=**'label'**,**x**=**'P'**,**data**=**da**[**da**[**'rainfall'**]>**150**])**

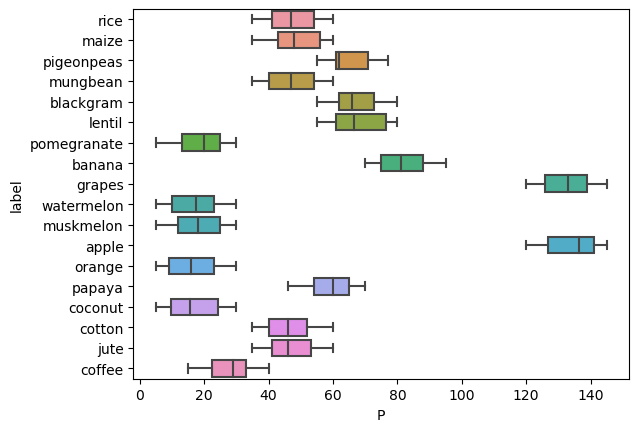
**INTERPRETATION:**

When it rains heavily (above 150 mm), Phosphorous levels in the environment become quite differentiable. This can be seen in water bodies, where soil and other sources release more phosphorous due to runoff. As the number of rain increases, so does the amount of phosphorous entering the water, creating a noticeable difference in phosphorous levels.

This difference can be seen in water clarity and aquatic plant growth. Heavy rain can also increase the rate of eutrophication in a lake, which is caused by an excessive amount of phosphorous in the water.

**BOXPLOT CONCERNING HUMIDITY**

sns**.**boxplot**(**y**=**'label'**,**x**=**'P'**,**data**=**da**[**da**[**'humidity'**]>**65**])**



**INTERPRETATION:**

The amount of phosphorous required for six crops grown in an area with humidity levels lower than 65 is relatively consistent, ranging from 14 to 25. This is regardless of the amount of rain expected over the next few weeks.

**DATA PRE-PROCESSING:**

#Data pre-processing

c**=**da**.**label**.**astype**(**'category'**)**

targets **=** dict**(**enumerate**(**c**.**cat**.**categories**))**

da**[**'target'**]=**c**.**cat**.**codes

y**=**da**.**target

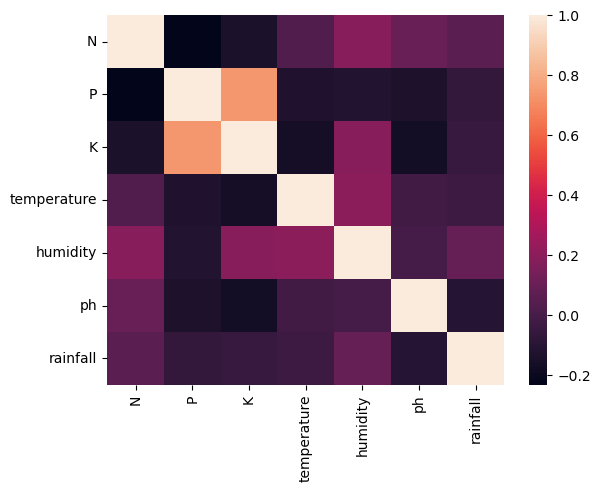
X**=**da**[[**'N'**,**'P'**,**'K'**,**'temperature'**,**'humidity'**,**'ph'**,**'rainfall'**]]**

**HEATMAP:**

#Heatmap

sns**.**heatmap**(**X**.**corr**())**

**INTERPRETATION:**

Correlation visualization is a useful tool for understanding the relationship between different features. In the case of phosphorous and potassium levels, a correlation visualization can show how the two are highly correlated, meaning that when one increases, the other increases too.

This can be useful for understanding the effects of changes on one on another and the effects of changes on one on the other.

**FEATURE SCALING:**

#Feature scaling

**from** sklearn**.**model\_selection **import** train\_test\_split

**from** sklearn**.**preprocessing **import** MinMaxScaler

X\_train**,** X\_test**,** y\_train**,** y\_test **=** train\_test\_split**(**X**,** y**,**random\_state**=**1**)**

scaler **=** MinMaxScaler**()**

X\_train\_scaled **=** scaler**.**fit\_transform**(**X\_train**)**

# we must apply the scaling to the test set as well that we are computing for the training set

X\_test\_scaled **=** scaler**.**transform**(**X\_test**)**

**EXPLANATION:**

Feature Scaling is an essential pre-processing step to ensure that all dataset features are brought to the same level of magnitude before applying any machine learning algorithm to the data. It helps to normalize the data within a specific range (e.g., 0 to 1).

As we saw in the code, two of our features, i.e., temperature and pH, are Gaussian distributed, scaling them between 0 and 1 with MinMaxScaler.

**DECISION TREE MODEL:**

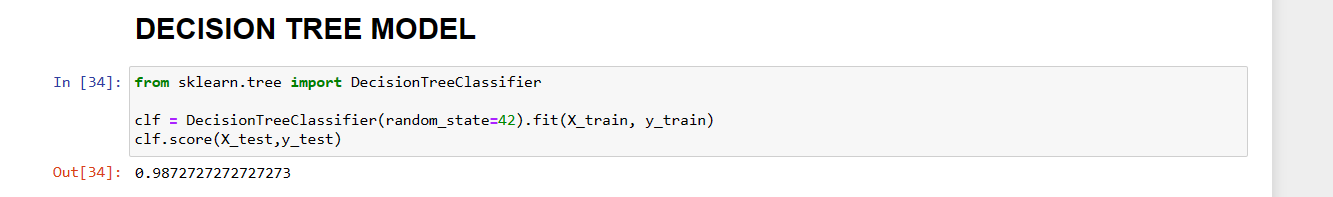
#dcision tree model

**from** sklearn**.**tree **import** DecisionTreeClassifier

clf **=** DecisionTreeClassifier**(**random\_state**=**42**).**fit**(**X\_train**,** y\_train**)**

clf**.**score**(**X\_test**,**y\_test**)**

**OUTPUT:**

****

**INTERPRETATION:**

The decision tree model, with an accuracy of 0.98, is a highly accurate model that can correctly classify the majority of the data points it is presented with. This high accuracy indicates a well-constructed and robust model and may result from careful data selection, training, parameter tuning, and testing.

**VISUALIZE FEATURE IMPORTANCE:**

#visualize feature importance

**import** numpy **as** np

plt**.**figure**(**figsize**=(**10**,**4**),** dpi**=**80**)**

c\_features **=** len**(**X\_train**.**columns**)**

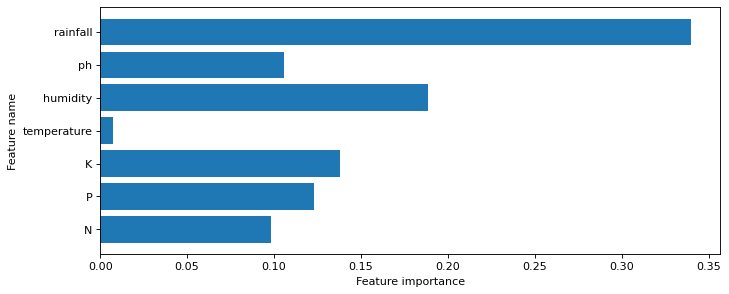
plt**.**barh**(**range**(**c\_features**),** clf**.**feature\_importances\_**)**

plt**.**xlabel**(**"Feature importance"**)**

plt**.**ylabel**(**"Feature name"**)**

plt**.**yticks**(**np**.**arange**(**c\_features**),** X\_train**.**columns**)**

plt**.**show**()**

**INTERPRETATION:**

When considering a decision tree, the feature importance measures how impact each feature has on the decision tree predictions. The algorithm determines the feature’s reputation based on how much information it offers and how many splits the decision tree can use. In the case of the rainfall feature, likely, the decision tree is more heavily relying on this feature because it has more predictive specific outcomes, such as crop yields or weather. Since rainfall can indicate how much moisture is in the environment, it is likely to be more relevant than other features when making predictions.