Page No.
Date: / /

6) Learning outcome:

Will be able to identify, visualize, process and relate the feature in the dataset.

7) Theory:

The iris dataset contains four features (length and width of sepals & petals) of 50 sample of three species of iris (iris setosa, iris virginica & iris versicolor). These measures were used to create a linear discriminant model to classify the species. The dataset is often used in data mining, classification and clustering examples and to test algorithms.

Description of features.

1) SepalLength cm - Sepal Length in centimeter.
2) Sepal Width cm - Sepal Width in centimeters.
3) PetalLength cm - Petal Length in centimeters.
4) Petal Width cm - Petal width in centimeters.
5) Species - species

Observations

1) In case of sepal length we can see that, the least sepal length counts around 1

Page No.
Date: / /

29 and the septal with max length that is around 7.7 is 1 in total.

2) In case of sepal width, the sepal with least width length are around 2 in total whereas the sepal with max long width are around 1 in total. Also on other hand if we look, so there are total 25+ sepals with width around 3.0 centimeters and there are around 10+ sepalse which have width varying from 2, 4, and 4.5.

3) To talk about petal length, the petal with least length are around 4 in total whereas the petal with max length are 4 in total too. On the other hand the max petal count has length of around 1.5 centimeters and the min petal counts has length of around 2.8 centimeters.

4) Among all the petal, the petal with least width are around 34 in total whereas the petal with max length are around 7 in total. On the other hand the petal with least count have width around 0.9 centimeters and the petal with max count have width around 34 in total.

Page No.
Date: / /

Outliers

Using boxplot for all features we can conclude that the feature sepal Length has no outlier because no value from it falls outside the quartile range.

In case of sepal width features, we can find that there are few values that are outside of inter quartile range so we can conclude that sepal width features have few outlier.

Similarly, for petal Length we can't find any values that lies outside the interquartile range so, petal Length also does not have any outliers.

In case of petal width feature, here too, we are unable to find any such values that lies outside the interquartile range, so this features also does not poses any outlier.

Conclusion:

Understood the relevance of features in the dataset using various plotting method and even found outliers for few features.