Understanding the dataset: IRIS DATASET

The Iris dataset contains 4 numerical attributes (4 measures) i.e., sepal length, sepal width, petal length and petal width and one predictive attribute i.e., species (Iris Setosa, Iris Versicolor, Iris Virginica). The 4 numerical attributes can be used to predict the species of the flower that has been classified into three species.

Chart, scatter chart

Description automatically generated

The above graph can be used to infer that the Iris can be classified using the petal length and petal width because the length and width of Iris setosa are much smaller than Iris versicolor and iris virginica. Then in the range of 1 to 1.7 cm petal width and 3 to 5 cm petal length the Iris versicolor can be classified and the Iris Setosa has the maximum range of petal width and length that is 1.5 to 2.5 cm and 5 to 7 cm.

Understanding the dataset: CENSUS DATASET

The data contains the following columns:

age

work class

fnlwgt: Number of people the entry represents

education

education-num: highest level of education represented numerically

marital status

occupation

relationship

race

sex

capital-gain

capital-loss

hours-per-week

native country

And the label income which is the predictive value that is income which helps us predict whether someone earns more than 50k or less.

Graphical user interface

Description automatically generated with low confidence

The above graph shows that most of the work class in the data is dominated my male sex. With only sectors such as private, local government and state government having close to 50 – 50 ratio between male and female works for all the sectors the male workforce is greater than the female workforce.

Understanding the dataset: CAR DATASET

The car dataset has the following attributes namely buying, maintenance, doors, persons, luggage boot (no. of luggage space in the boot) and safety. The dataset also has a predictive column that helps us in predicting whether the car is **acceptable or not** based on the above attributes (basically divided into two parts price (buying and maintaining) and comfort (door, persons, luggage boot)).

Chart, bar chart, waterfall chart

Description automatically generated

The above graph helps us in inferring that the cars that have a low safety plays an important role in deciding whether the condition of the car is acceptable or not. As we can see that in any of the other classes above that is acceptable, good and vgood no cars with low safety are placed meaning are with medium or high safety qualify to be in acceptable good and vgood columns.



The pie chart above helps us in inferring that the cars that are unacceptable have no strong correlation with the buying and maintenance columns as the pir charts formed are equally distributed.

Understanding the dataset: MUSHROOM DATASET

In the Mushroom dataset there are 22 categorical data. There are divided based on cap, gill, stalk, odor, habitat, veil, ring etc. The class of the dataset is edible or poisonous we can use the 22 features of the dataset to predict whether the mushroom is edible or not (poisonous) .

Graphical user interface, application

Description automatically generated

The graph uses the feature odor to infer that odor (almond=a, anise=l, creosote=c, fishy=y, foul=f, musty=m, none=n, pungent=p, spicy=s) The mushrooms having creosote, foul, musty, pungent, spicy and fishy odor are 100% poisonous.