# Assignment **1**

# Implementation of ID3 Algorithm by using Various Discretization Techniques

## Iris Dataset

The iris dataset contains a set of 150 records under 5 attributes - Petal Length, Petal Width, Sepal Length, Sepal width and Class (Species). The data set consists of 50 samples from each of three species of Iris (Iris Setosa, Iris virginica, and Iris versicolor). Four features were measured from each sample: the length and the width of the sepals and petals, in centimeters.

## Discretization Techniques

In this assignment we have used two discretization techniques namely

1. Naïve Discretization Technique
2. CACC Discretization Technique

### Naïve Discretization Technique

In this technique we will discretize the data by rounding the continuous values to nearest integers. This is done using round ( ).

### CACC Discretization Technique

CACC Discretization stands for Class-Attribute Contingency Coefficient Discretization. Using CACC Algorithm we have discretized the data, we have used a function CACC for discretization purpose .This function takes iris dataset as input and return three objects as output:

1) Disc Data

2) Disc Values

3) Disc Schema

Disc data is MxN matrix containing the discretized data of the iris dataset using CACC algorithm.

Disc Values is an Nx1 cell containing the possible values of each feature.

Disc Schema is an N-1x1 cell containing the boundaries for each feature.

## ID3 Algorithm

ID3 algorithm stands for Iterative Dichotomiser 3 algorithm. Using ID3 algorithm we have generated a decision tree. Accuracies are calculated by testing the test data against the training data with the help of the decision tree.

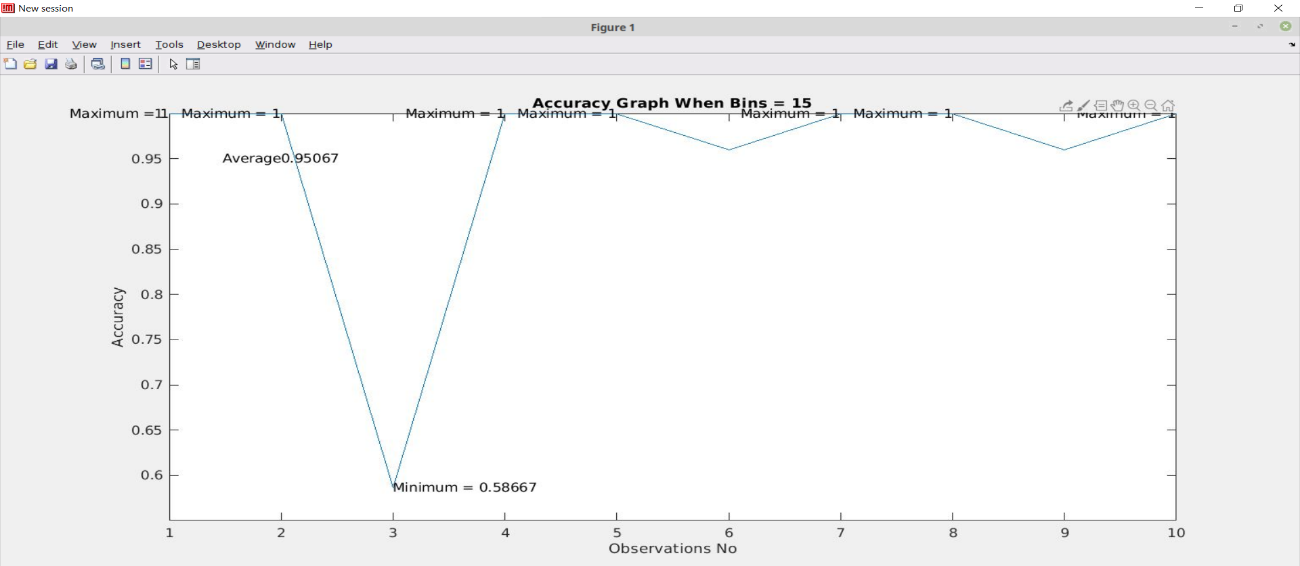
## Implementation

The project is developed and executed in MATLAB software. Initially the Iris dataset is loaded using fscan( ). Then the user is prompted to select either Naïve discretization approach or CACC discretization approach. Upon selecting one approach the entire dataset is divided into test and training data by using randperm( ). The randperm( ) returns a random row vector from a given data. With the obtained training data the ID3 algorithm constructs a decision tree. Now by testing the test records using the constructed decision tree the accuracy value is obtained. This testing process is repeated for 10 times with different test and training records. All the obtained accuracies are plotted as a graph using plot( ). The average of all the accuracy values is also displayed in the graph as shown in the figure.

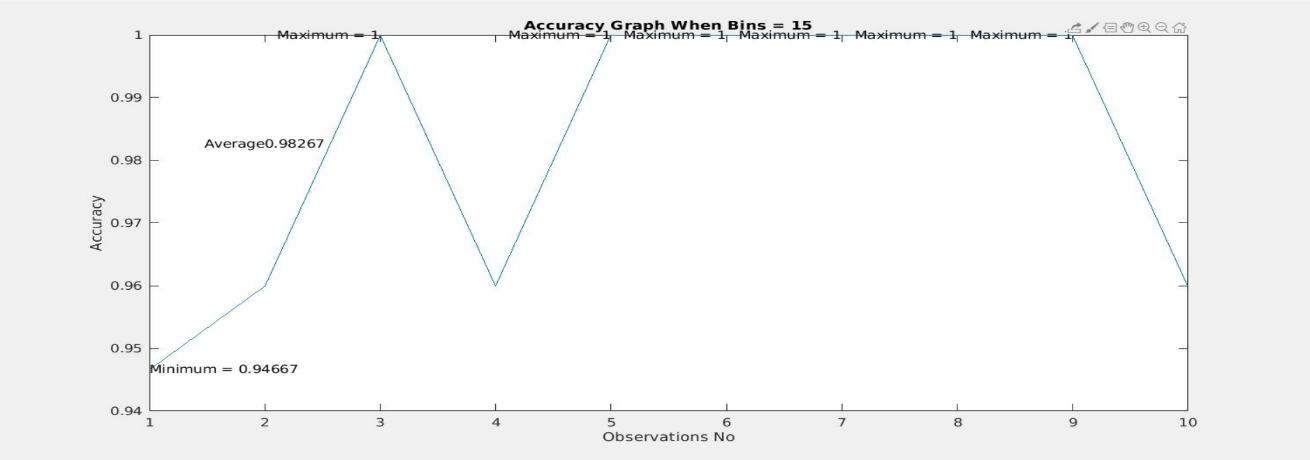
## Observations:

When we the discretized data using round function and sent input to the learning model with random training data and testing data we observed an average accuracy of 0.9787 for a total of 10 runs.

Accuracy graphs at two instances for ID3 using Naïve approach:

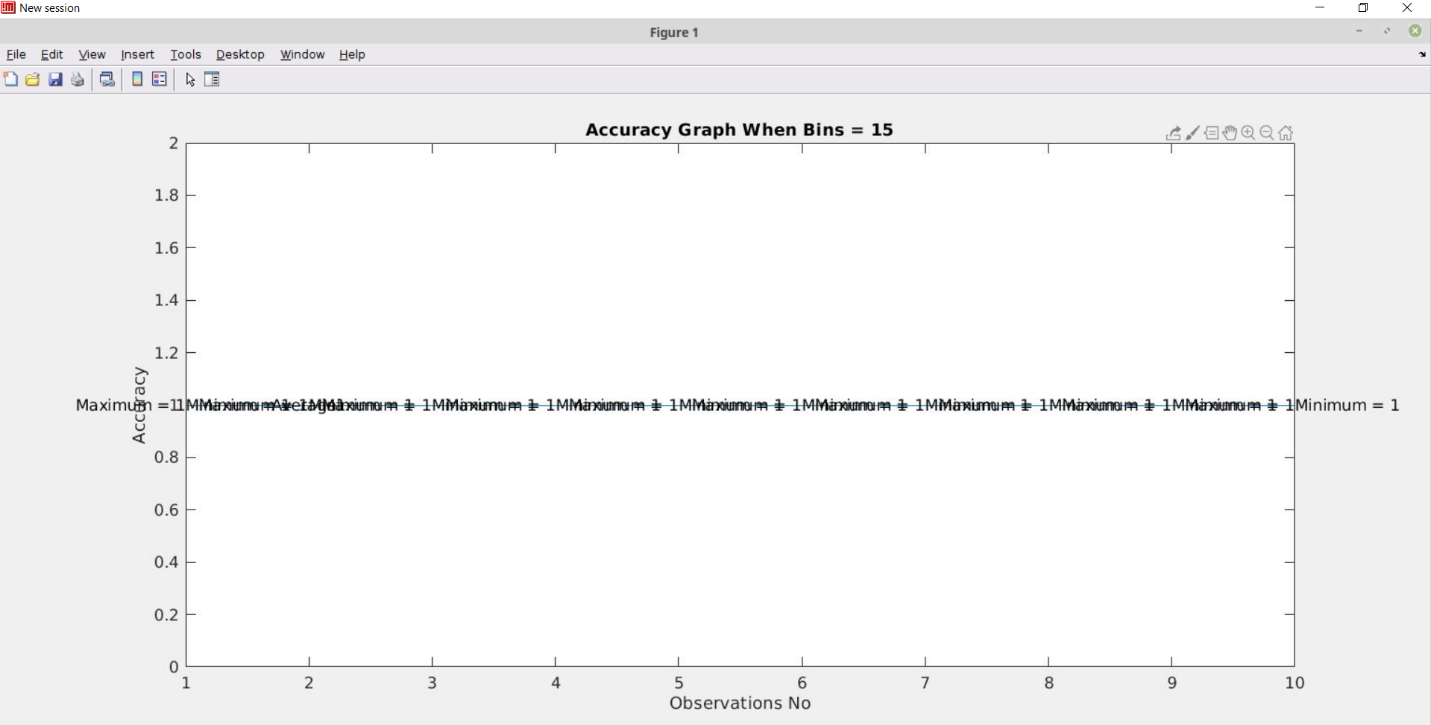


Screen shot for ID3 implementation with naïve approach, Instance 1:



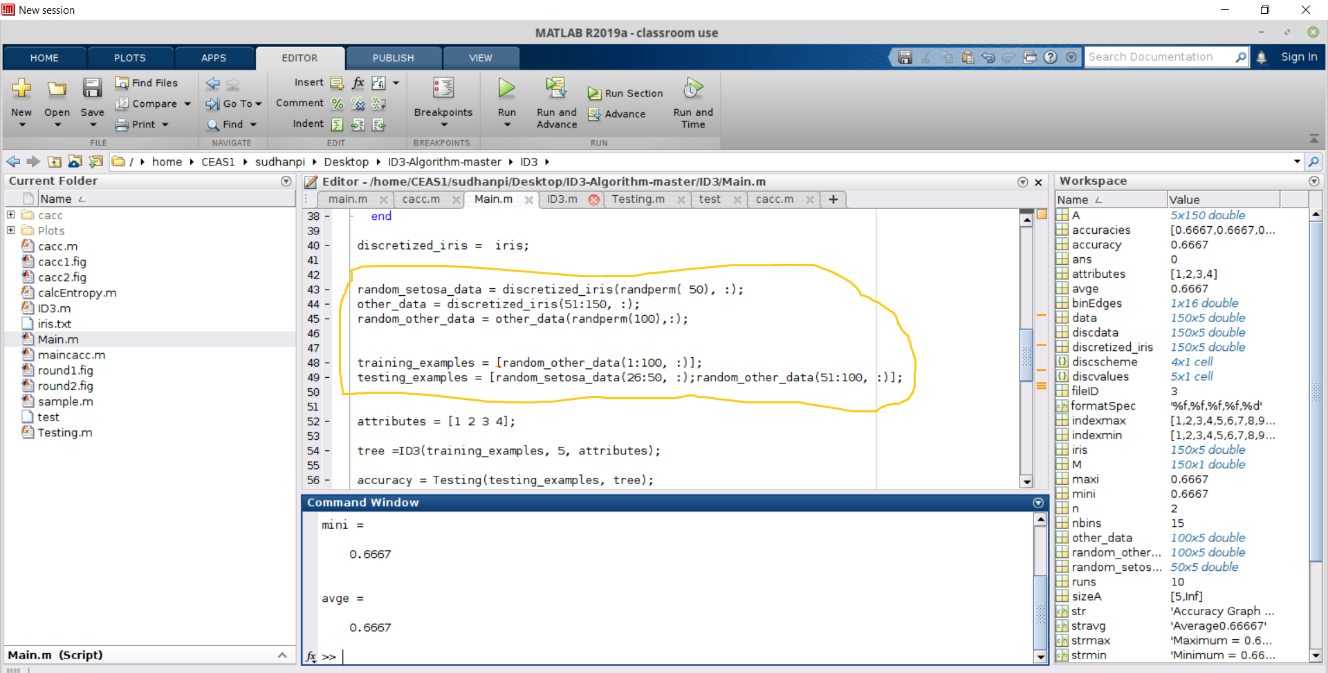
Screen shot for ID3 implementation with naïve approach, Instance 2:

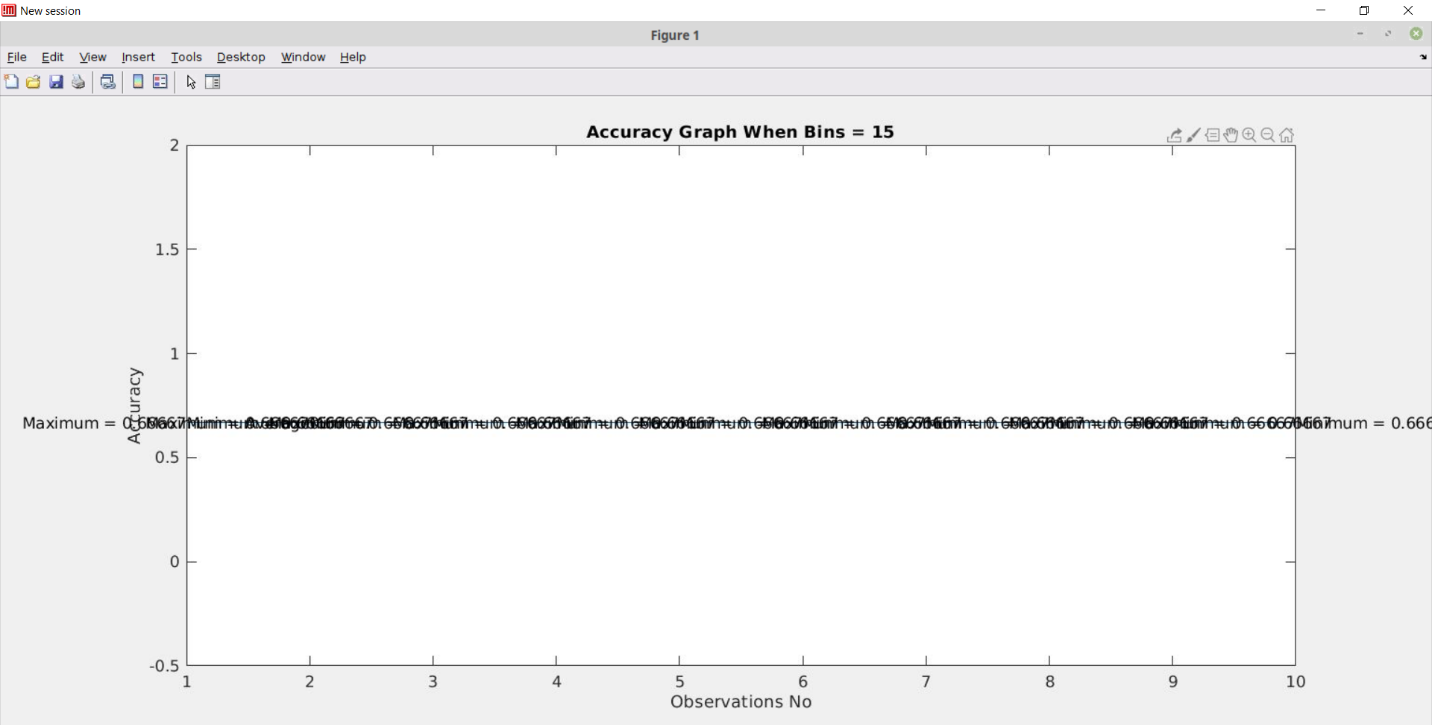
When the data gets discretized using CACC alogrithm we are getting 100 % accuracy .



Screen shot for ID3 implementation with CACC approach, Instance 1:

To test our code we have created a case to train the learning model with only two classes of flowers and sent the test data with all 150 samples, theatrically my model has to detect only 100 samples out of 150, so the accuracy would be 100/150=2/3\*100=66.66 and below is the snap of code change for this case.





ID3 Implementation with Cacc discretization technique screen shot with 66.7 % accuracy

iris dataset discretization using naïve approach:  
  
     5     4     1     0     1  
     5     3     1     0     1  
     5     3     1     0     1  
     5     3     2     0     1  
     5     4     1     0     1  
     5     4     2     0     1  
     5     3     1     0     1  
     5     3     2     0     1  
     4     3     1     0     1  
     5     3     2     0     1  
     5     4     2     0     1  
     5     3     2     0     1  
     5     3     1     0     1  
     4     3     1     0     1  
     6     4     1     0     1  
     6     4     2     0     1  
     5     4     1     0     1  
     5     4     1     0     1  
     6     4     2     0     1  
     5     4     2     0     1  
     5     3     2     0     1  
     5     4     2     0     1  
     5     4     1     0     1  
     5     3     2     1     1  
     5     3     2     0     1  
     5     3     2     0     1  
     5     3     2     0     1  
     5     4     2     0     1  
     5     3     1     0     1  
     5     3     2     0     1  
     5     3     2     0     1  
     5     3     2     0     1  
     5     4     2     0     1  
     6     4     1     0     1  
     5     3     2     0     1  
     5     3     1     0     1  
     6     4     1     0     1  
     5     3     2     0     1  
     4     3     1     0     1  
     5     3     2     0     1  
     5     4     1     0     1  
     5     2     1     0     1  
     4     3     1     0     1  
     5     4     2     1     1  
     5     4     2     0     1  
     5     3     1     0     1  
     5     4     2     0     1  
     5     3     1     0     1  
     5     4     2     0     1  
     5     3     1     0     1  
     7     3     5     1     2  
     6     3     5     2     2  
     7     3     5     2     2  
     6     2     4     1     2  
     7     3     5     2     2  
     6     3     5     1     2  
     6     3     5     2     2  
     5     2     3     1     2  
     7     3     5     1     2  
     5     3     4     1     2  
     5     2     4     1     2  
     6     3     4     2     2  
     6     2     4     1     2  
     6     3     5     1     2  
     6     3     4     1     2  
     7     3     4     1     2  
     6     3     5     2     2  
     6     3     4     1     2  
     6     2     5     2     2  
     6     3     4     1     2  
     6     3     5     2     2  
     6     3     4     1     2  
     6     3     5     2     2  
     6     3     5     1     2  
     6     3     4     1     2  
     7     3     4     1     2  
     7     3     5     1     2  
     7     3     5     2     2  
     6     3     5     2     2  
     6     3     4     1     2  
     6     2     4     1     2  
     6     2     4     1     2  
     6     3     4     1     2  
     6     3     5     2     2  
     5     3     5     2     2  
     6     3     5     2     2  
     7     3     5     2     2  
     6     2     4     1     2  
     6     3     4     1     2  
     6     3     4     1     2  
     6     3     4     1     2  
     6     3     5     1     2  
     6     3     4     1     2  
     5     2     3     1     2  
     6     3     4     1     2  
     6     3     4     1     2  
     6     3     4     1     2  
     6     3     4     1     2  
     5     3     3     1     2  
     6     3     4     1     2  
     6     3     6     3     3  
     6     3     5     2     3  
     7     3     6     2     3  
     6     3     6     2     3  
     7     3     6     2     3  
     8     3     7     2     3  
     5     3     5     2     3  
     7     3     6     2     3  
     7     3     6     2     3  
     7     4     6     3     3  
     7     3     5     2     3  
     6     3     5     2     3  
     7     3     6     2     3  
     6     3     5     2     3  
     6     3     5     2     3  
     6     3     5     2     3  
     7     3     6     2     3  
     8     4     7     2     3  
     8     3     7     2     3  
     6     2     5     2     3  
     7     3     6     2     3  
     6     3     5     2     3  
     8     3     7     2     3  
     6     3     5     2     3  
     7     3     6     2     3  
     7     3     6     2     3  
     6     3     5     2     3  
     6     3     5     2     3  
     6     3     6     2     3  
     7     3     6     2     3  
     7     3     6     2     3  
     8     4     6     2     3  
     6     3     6     2     3  
     6     3     5     2     3  
     6     3     6     1     3  
     8     3     6     2     3  
     6     3     6     2     3  
     6     3     6     2     3  
     6     3     5     2     3  
     7     3     5     2     3  
     7     3     6     2     3  
     7     3     5     2     3  
     6     3     5     2     3  
     7     3     6     2     3  
     7     3     6     3     3  
     7     3     5     2     3  
     6     3     5     2     3  
     7     3     5     2     3  
     6     3     5     2     3  
     6     3     5     2     3

iris dataset data after CACC discretization technique:  
  
     3     4     1     1     1  
     2     3     1     1     1  
     2     3     1     1     1  
     2     3     1     1     1  
     2     4     1     1     1  
     3     4     1     1     1  
     2     4     1     1     1  
     2     4     1     1     1  
     2     2     1     1     1  
     2     3     1     1     1  
     3     4     1     1     1  
     2     4     1     1     1  
     2     3     1     1     1  
     1     3     1     1     1  
     4     4     1     1     1  
     4     5     1     1     1  
     3     4     1     1     1  
     3     4     1     1     1  
     4     4     1     1     1  
     3     4     1     1     1  
     3     4     1     1     1  
     3     4     1     1     1  
     2     4     1     1     1  
     3     3     1     1     1  
     2     4     1     1     1  
     2     3     1     1     1  
     2     4     1     1     1  
     3     4     1     1     1  
     3     4     1     1     1  
     2     3     1     1     1  
     2     3     1     1     1  
     3     4     1     1     1  
     3     5     1     1     1  
     4     5     1     1     1  
     2     3     1     1     1  
     2     3     1     1     1  
     4     4     1     1     1  
     2     3     1     1     1  
     2     3     1     1     1  
     3     4     1     1     1  
     2     4     1     1     1  
     2     1     1     1     1  
     2     3     1     1     1  
     2     4     1     1     1  
     3     4     1     1     1  
     2     3     1     1     1  
     3     4     1     1     1  
     2     3     1     1     1  
     3     4     1     1     1  
     2     3     1     1     1  
     5     3     3     2     2  
     5     3     3     2     2  
     5     3     3     2     2  
     4     1     2     2     2  
     5     2     3     2     2  
     4     2     3     2     2  
     5     3     3     2     2  
     2     1     2     2     2  
     5     2     3     2     2  
     3     2     2     2     2  
     2     1     2     2     2  
     4     3     3     2     2  
     4     1     2     2     2  
     4     2     3     2     2  
     4     2     2     2     2  
     5     3     3     2     2  
     4     3     3     2     2  
     4     2     3     2     2  
     5     1     3     2     2  
     4     2     2     2     2  
     4     3     3     3     2  
     4     2     2     2     2  
     5     2     3     2     2  
     4     2     3     2     2  
     5     2     3     2     2  
     5     3     3     2     2  
     5     2     3     2     2  
     5     3     3     2     2  
     4     2     3     2     2  
     4     2     2     2     2  
     4     1     2     2     2  
     4     1     2     2     2  
     4     2     2     2     2  
     4     2     4     2     2  
     3     3     3     2     2  
     4     4     3     2     2  
     5     3     3     2     2  
     5     1     3     2     2  
     4     3     3     2     2  
     4     2     2     2     2  
     4     2     3     2     2  
     4     3     3     2     2  
     4     2     2     2     2  
     2     1     2     2     2  
     4     2     3     2     2  
     4     3     3     2     2  
     4     2     3     2     2  
     5     2     3     2     2  
     3     2     2     2     2  
     4     2     3     2     2  
     5     3     4     3     3  
     4     2     4     3     3  
     5     3     4     3     3  
     5     2     4     3     3  
     5     3     4     3     3  
     5     3     4     3     3  
     2     2     3     2     3  
     5     2     4     3     3  
     5     2     4     3     3  
     5     4     4     3     3  
     5     3     4     3     3  
     5     2     4     3     3  
     5     3     4     3     3  
     4     2     3     3     3  
     4     2     4     3     3  
     5     3     4     3     3  
     5     3     4     3     3  
     5     4     4     3     3  
     5     2     4     3     3  
     4     1     3     2     3  
     5     3     4     3     3  
     4     2     3     3     3  
     5     2     4     3     3  
     5     2     3     3     3  
     5     3     4     3     3  
     5     3     4     3     3  
     5     2     3     3     3  
     4     3     3     3     3  
     5     2     4     3     3  
     5     3     4     2     3  
     5     2     4     3     3  
     5     4     4     3     3  
     5     2     4     3     3  
     5     2     4     2     3  
     4     2     4     2     3  
     5     3     4     3     3  
     5     4     4     3     3  
     5     3     4     3     3  
     4     3     3     3     3  
     5     3     4     3     3  
     5     3     4     3     3  
     5     3     4     3     3  
     4     2     4     3     3  
     5     3     4     3     3  
     5     3     4     3     3  
     5     3     4     3     3  
     5     2     3     3     3  
     5     3     4     3     3  
     5     4     4     3     3  
     4     3     4     3     3

### Team Members:

Kiran Kumar Dogiparthi

Prudhvi Sudhana

Veerendra Kumar Pardhu Balisetty

References:

<https://www.mathworks.com/matlabcentral/fileexchange/41740-discretization-methods-class-attribute-contingency-coefficient-cacc-matlab>

https://github.com/Anish1998/ID3-Algorithm-Implementation.git

Used Math works website to know information about various functions like fscanf,randperm,discretize, struct