



Mahavir Education Trust's
**SHAH & ANCHOR KUTCHHI ENGINEERING
COLLEGE**
Chembur, Mumbai - 400 088
UG Program in Cyber Security

Experiment Number: 5					
Date of Performance:					
Date of Submission:					
Program Execution/ formation/ correction/ ethical practices (07)	Documentation (02)	Timely Submission (03)	Viva Answer to sample questions (03)	Experiment Total (15)	Sign

Experiment 5

Aim: Implementation of Data Discretization (any one) & Visualization (any one).

Lab outcomes: CSL 503.2: Implement data mining algorithms like classification.

Problem Statement: Implement Data Discretization

Theory:

Data Discretization: Data discretization refers to a method of converting a huge number of data values into smaller ones so that the evaluation and management of data become easy. In other words, data discretization is a method of converting attributes values of continuous data into a finite set of intervals with minimum data loss. There are two forms of data discretization first is supervised discretization, and the second is unsupervised discretization. Supervised discretization refers to a method in which the class data is used. Unsupervised discretization refers to a method depending upon the way which operation proceeds. It means it works on the top-down splitting strategy and bottom-up merging strategy.

Some Famous techniques of data discretization

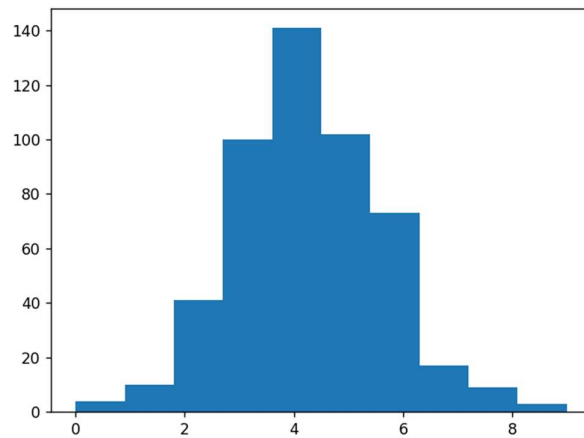
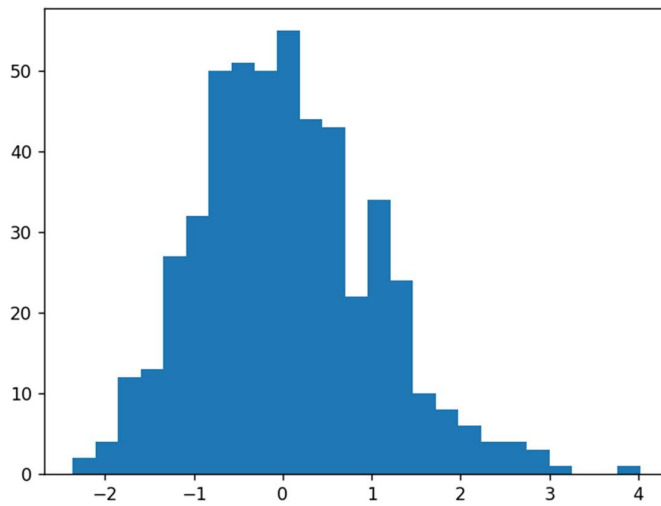
- ❑ Histogram analysis
- ❑ Binning
- ❑ Cluster Analysis
- ❑ Data discretization using decision tree analysis
- ❑ Data discretization using correlation analysis

Data Visualization: Data Visualization is the representation of data within the frame of the chart, chart, picture, etc. These are made as the visual representation of information. It makes a difference a parcel to choose makes to see the analytics displayed outwardly, so they can work in like manner and can get a handle on the troublesome concept or recognize modern designs.

Program Listing and Output:

```
from numpy.random import randn
from sklearn.preprocessing import KBinsDiscretizer
from matplotlib import pyplot
data=randn(500)
pyplot.hist(data,bins=25)
pyplot.show()
data=data.reshape((len(data),1))
kbins=KBinsDiscretizer(n_bins=10,encode='ordinal',strategy='uniform')
data_trans=kbins.fit_transform(data)
print(data_trans[:10,:])
pyplot.hist(data_trans, bins=10)
pyplot.show()
```

Output:-



```
[[5.]
 [5.]
 [5.]
 [6.]
 [3.]
 [5.]
 [5.]
 [5.]
 [2.]
 [5.]]
```

Conclusion: Here we implemented Data Discretization using python.