Final Year B.Tech. (CSE) – I : [2022-23]

5CS462 : PE5 Data Mining Lab

PRN/ Roll No: 2019BTECS00113

Full name: Sanket Rajaram Mote

Batch: B8

Assignment No. 2

Title: Data Analysis Tool (GUI)

Objective:

To calculate Chi- Square

2. To Calculate Pearson coefficient and covariance

3. Implement Min-max Normalization

4. To Implement Z-Score normalisation

Introduction & Theory:

1.Chi-Square Test :

A chi-square test is used to help determine if observed results are in line with expected results, and to rule out that observations are due to chance

2. Correlation coefficient :

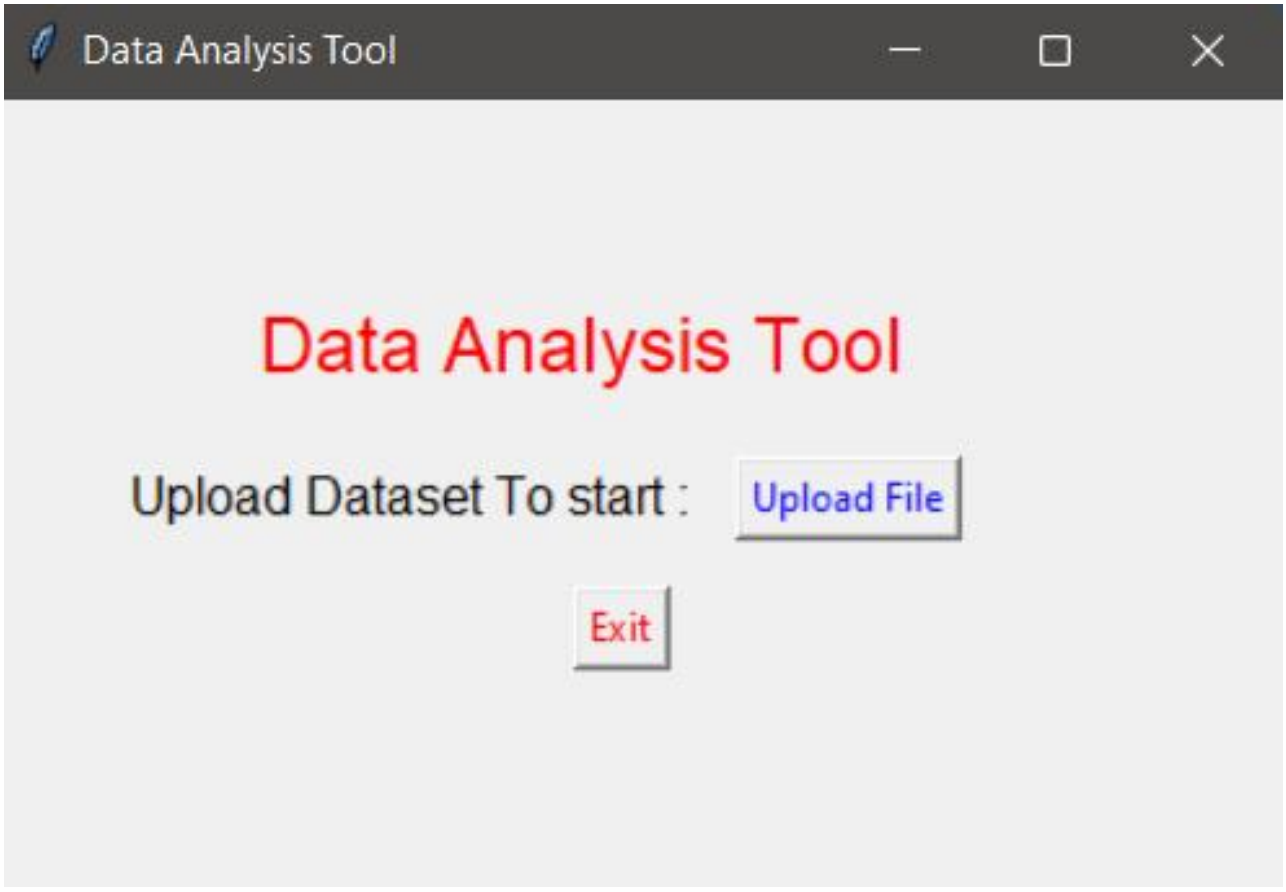
It measures the extent to which the value of one variable changes with another.

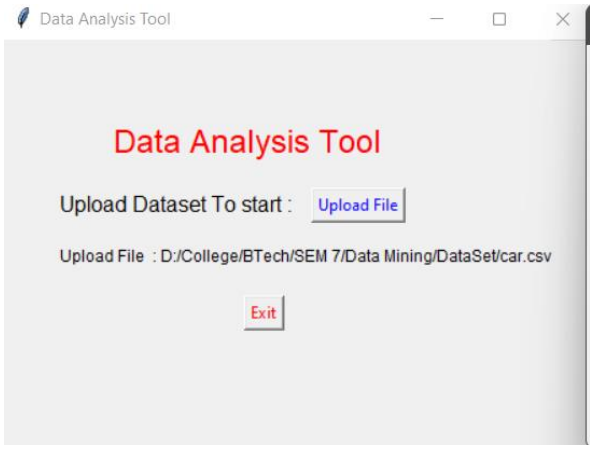
3.Min-max normalization : Min Max is a data normalization technique like Z score, decimal scaling, and normalization with standard deviation.

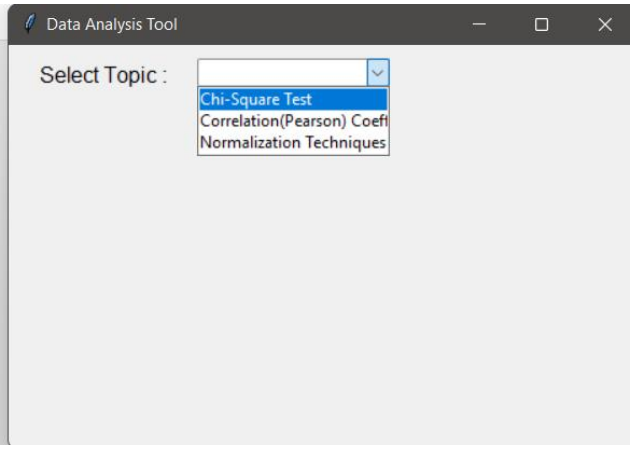
4. Z-Score normalization :

Z-Score value is to understand how far the data point is from the mean. Technically, it measures the standard deviations below or above the mean.

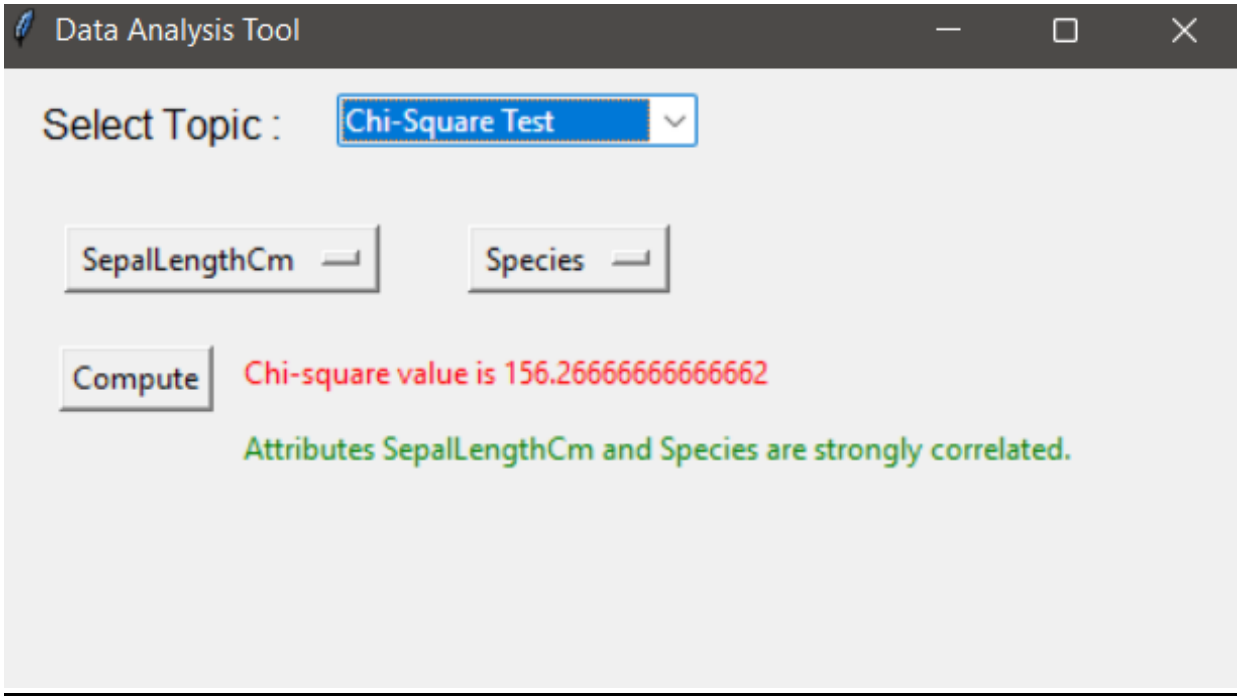
Screen Shots :

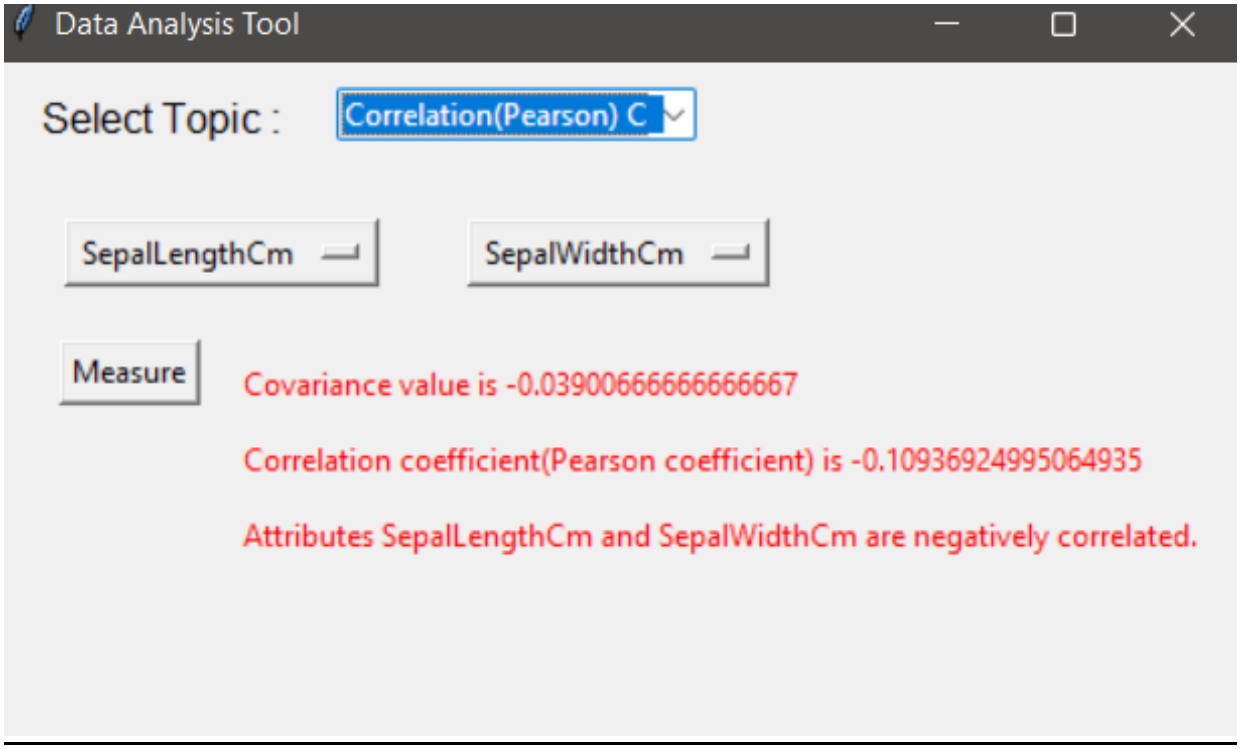


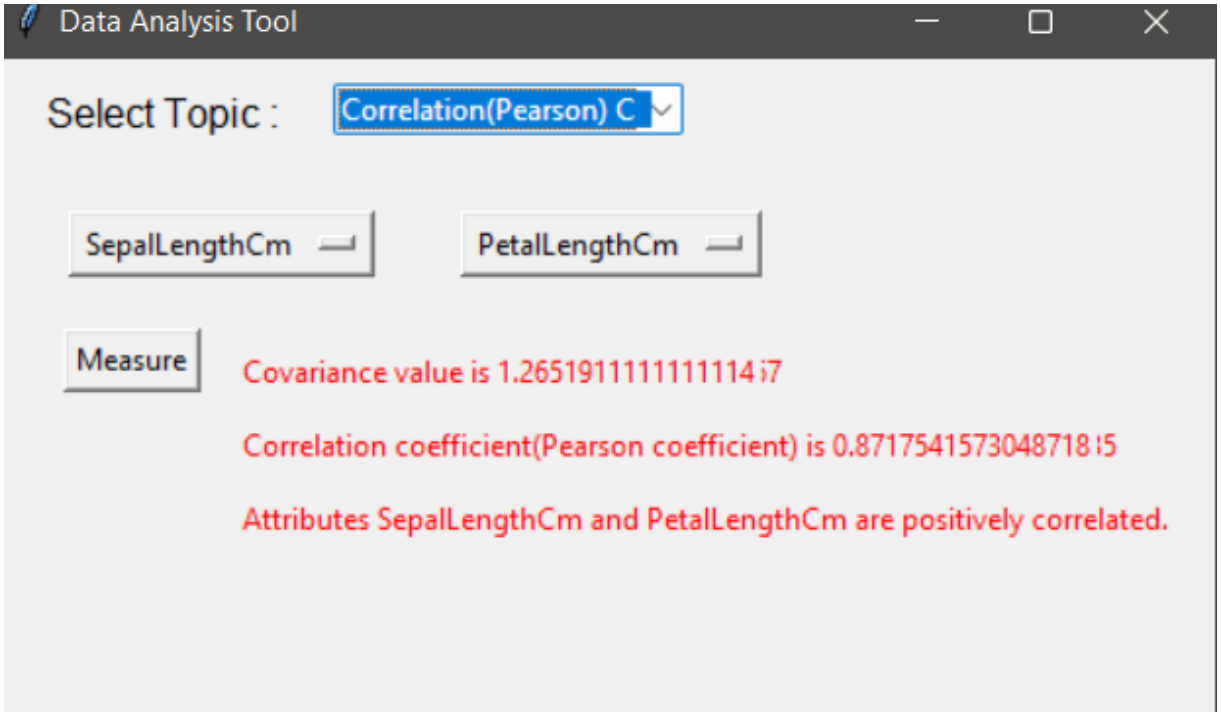


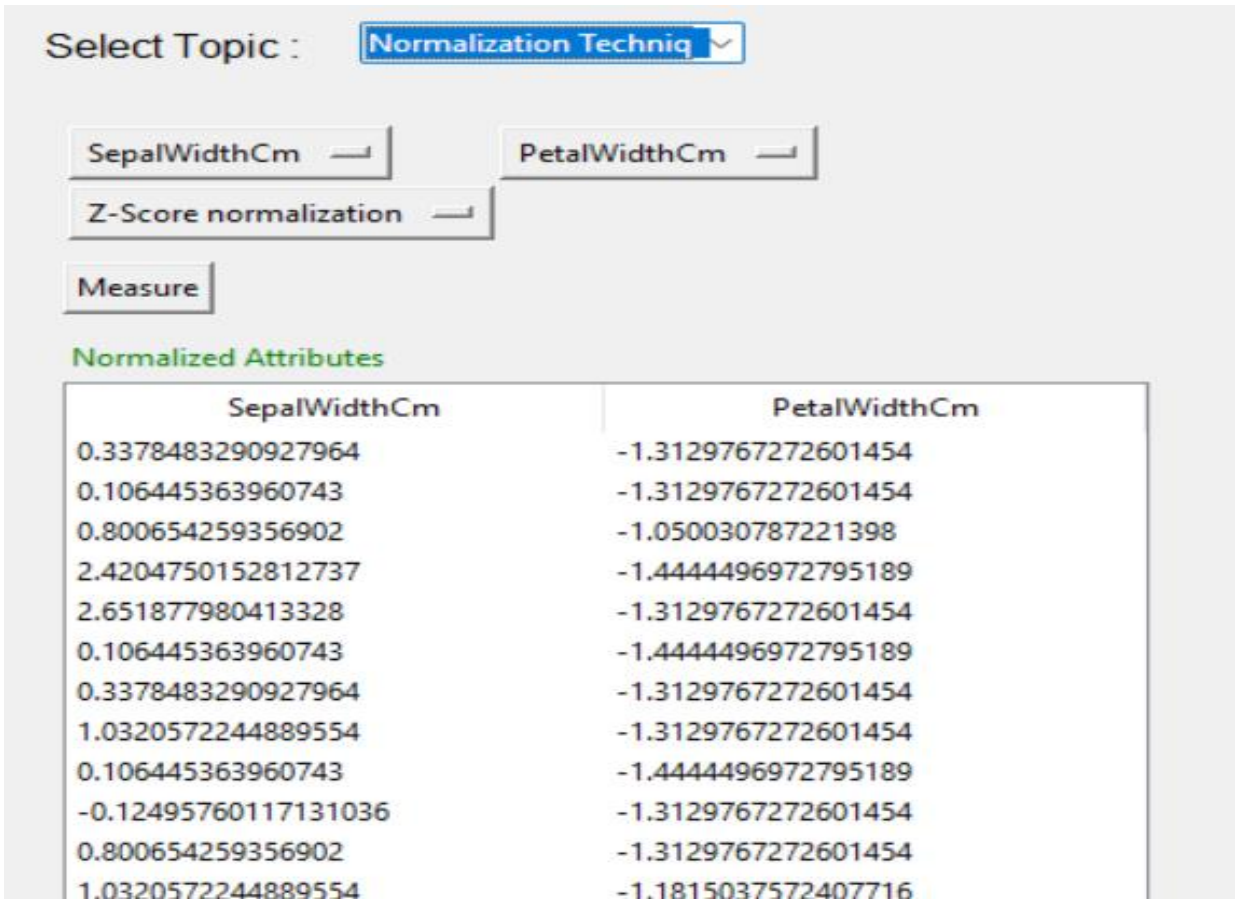


Iris-setosa	Iris-versicolor	Iris-virginica	Total
1	0	0	1
3	0	0	3
1	0	0	1
4	0	0	4
2	0	0	2
5	0	0	5


Data Analysis Tool
Select Topic :
Chi-Square Test
SepalLengthCm
Species
Compute
Chi-square value is 156.26666666666662
Attributes SepalLengthCm and Species are strongly correlated.


Data Analysis Tool
Select Topic :
Correlation(Pearson) C
SepalLengthCm
SepalWidthCm
Measure
Covariance value is -0.03900666666666667
Correlation coefficient(Pearson coefficient) is -0.10936924995064935
Attributes SepalLengthCm and SepalWidthCm are negatively correlated.


Data Analysis Tool
Select Topic :
Correlation(Pearson) C
SepalLengthCm
PetalLengthCm
Measure
Covariance value is 1.26519111111111114 i7
Correlation coefficient(Pearson coefficient) is 0.8717541573048718 i5
Attributes SepalLengthCm and PetalLengthCm are positively correlated.


Select Topic :
Normalization Techniq
SepalWidthCm
PetalWidthCm
Z-Score normalization
Measure
Normalized Attributes
SepalWidthCm
PetalWidthCm
0.3378483290927964
-1.3129767272601454
0.106445363960743
-1.3129767272601454
0.800654259356902
-1.050030787221398
2.4204750152812737
-1.4444496972795189
2.651877980413328
-1.3129767272601454
0.106445363960743
-1.4444496972795189
0.3378483290927964
-1.3129767272601454
1.0320572244889554
-1.3129767272601454
0.106445363960743
-1.4444496972795189
-0.12495760117131036
-1.3129767272601454
0.800654259356902
-1.3129767272601454
1.0320572244889554
-1.1815037572407716

Conclusion:

Successfully implemented chi square test , Correlation coefficient , Covariance, Different Type of Normalisation using Gui

References:

https://www.geeksforgeeks.org/data-normalization-in-data-mining/

https://www.kaggle.com/code/dehaozhang/chi-squared-test-on-attrition/notebook

Dataset:

https://archive.ics.uci.edu/ml/datasets/Iris

https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29