Department of Computer Science & Engineering

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

5CS462 : PE5 - Data Mining Lab

Assignment No. 3

Use/extend the data analysis tool (GUI) developed in Assignment No. 2 to perform the following pre-processing task :

1. Correlation analysis - Chi-Square Test
   a. User should be able to choose any two attributes.
   b. Display the contingency table
   c. Show the chi-square value and conclusion whether the selected attributes are correlated or not.
2. Correlation analysis – Correlation coefficient (Pearson coefficient) & Covariance
   a) User should be able to choose any two attributes.
   b) Show the calculated values and conclusion whether the selected attributes are correlated or not.
3. Normalization using following techniques :
   a. Min-max normalization
   b. Z-Score normalization
   c. Normalization by decimal scaling

   i. User should be able to choose any attributes.
   ii. Show the calculated values in tabulated form.
   iii. Show the scatter plot for normalized attributes

Note :

- **Don't use direct / readymade functions from library. Do basic python coding.** You may compare / check the results with std. libraries.
- Use IRIS, Breast Cancer Wisconsin data sets *OR any other data sets* from UCI Machine Learning Repository (http://archive.ics.uci.edu/ml/).

Deadline : 21/08/2022

Dr. B. F. Momin
Course Teacher