COLLEGE OF ENGINEERING AND MANAGEMENT PUNNAPRA

Under CAPE | Government of Kerala

Issue No. / Date: 01/26.07.2019 **Revn. No. / Date:00**	**TUTORIAL -3**	**CEMP/ISO/UG/CSE/653 I**

Programme: B.TECH		**Branch: COMPUTER SCIENCE & ENGINEERING**
Semester: S8		**Academic Year: 2022-2023**
Subject Code: CST 466	**L-T-P- Credits: 2-1-0-3**	**Subject Name: DATA MINING**
Date:	**Hour:**	**Topic: Data Preprocessing**

Objective

To understand the basic ideas of Steps in preprocessing

Questions

1. Suppose that the data for analysis includes the attribute age. The age values for the data tuples are (in increasing order) 13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70.
   (a) What is the mean of the data? What is the median?
   (b) Use smoothing by bin means to smooth the data, using a bin depth of 3. Illustrate your steps. Comment on the effect of this technique for the given data.
   (c) How might you determine outliers in the data?
   (d) What other methods are there for data smoothing?
   (e) Use min-max normalization to transform the value 35 for age onto the range [0:0;1:0].
   (f) Use z-score normalization to transform the value 35 for age, where the standard deviation of age is 12.94 years.
   (g) Use normalization by decimal scaling to transform the value 35 for age.
   (h) Comment on which method of normalization you would prefer to use for the given data, giving reasons as to why.
2. Suppose a hospital tested the age and body fat data for 18 randomly selected adults with the following result:

age	23	23	27	27	39	41	47	49	50
%fat	9.5	26.5	7.8	17.8	31.4	25.9	27.4	27.2	31.2
age	52	54	54	56	57	58	58	60	61
%fat	34.6	42.5	28.8	33.4	30.2	34.1	32.9	41.2	35.7

   a. Calculate the mean, median, and standard deviation of age and %fat.
   b. Normalize the two variables based on z-score normalization.
   c. Calculate the correlation coefficient (Pearson's product moment coefficient). Are these two variables positively or negatively correlated?

1. 1. Dunham M H, "Data Mining: Introductory and Advanced Topics", Pearson Education, New Delhi, 2003.