



আন্তর্জাতিক ইসলামী বিশ্ববিদ্যালয় চট্টগ্রাম
الجامعة الإسلامية شيتاغونغ
International Islamic University Chittagong

LAB REPORT

Department : Computer Science and Engineering
Course Title : Machine Learning and Data Mining Lab
Course Code : CSE-4878

No of Experiment : 01

Name of Experiment :

Submitted To

Mrs. Subrina Akter

Submitted By

Name : Tasnimul Jannat Niba, Umme Kulsum Neha (Group-10)

ID No : C191242, C191272

Semester : 8th

Section : 8BF.

Date of Issue :

Date of Submission : 31/01/2023

Remark



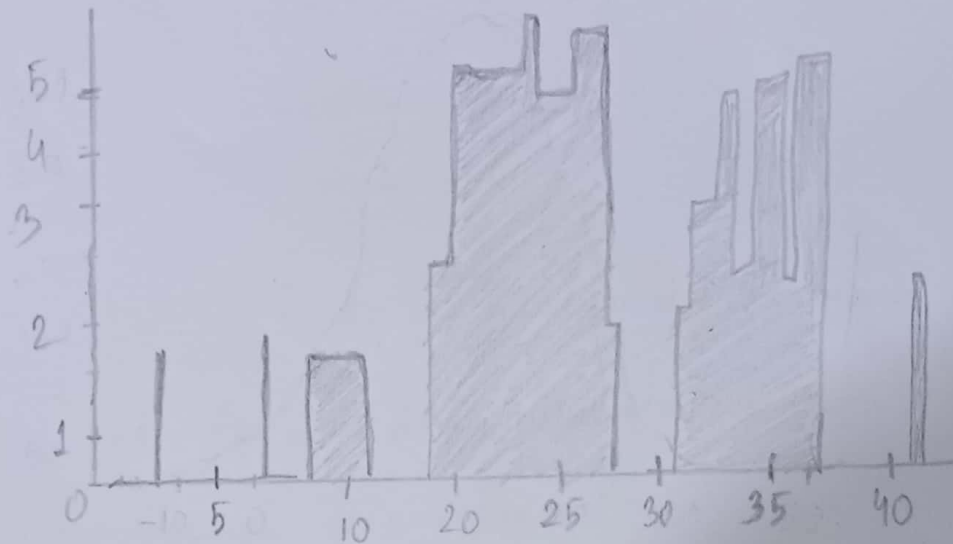
Dataset: We have chosen the dataset of Students Mental Health.

Dataset Information: This dataset has 11 features. The features are Timestamp, gender, Age, Course, year, CGPA, Marital status, Depression condition, Anxiety, panic attack & special treatment. This dataset is collected from a survey in order to examine students current academic situation and mental health.

Characteristics of the Dataset:

☐ Normally distributed or not: In order to be considered a normal distribution, a dataset (when graphed) must follow a bell-shaped symmetrical curve centered around the mean.

In this dataset, if we take the Age attribute of students we get,



• Age Col: From the Age col we get the following mean, median & mode,

$$\text{mean} = 21.55$$

$$\text{median} = 19.0$$

$$\text{mode} = 18.0$$

We know, Graph is normally distributed when the value of mean, median & mode is equal.

As here, $\text{mean} \neq \text{median} \neq \text{mode}$

so, it is not normally distributed.

• CGPA Col: From the CGPA col we get the following result,

$$\text{mean} = 3.4945545545545$$

$$\text{median} = 3.49$$

$$\text{mode} = 4.0$$

▣ Find Skewness: As we mentioned the mean, median & mode of Age & CGPA attributes, we can find skewness through this information,

For Age,

$$\text{mode} < \text{median} < \text{mean}$$

So, this is positively skewed.

For CGPA,

$$\text{mean} < \text{median} < \text{mode}$$

Thus, it is negatively skewed.

▣ Find Missing value: We found that there exists some missing value. The Age attribute has 1 missing value & the CGPA attribute has 3 missing value.

Age	1
CGPA	3

▣ Fill Missing value: As we know, for ordinal, Interval and ratio data (if skewed) then we have to consider median value to consider the missing value. Thus we filled the missing value with Median value.

▣ Find Variance: Variance is a measurement used to determine how far each number is from the mean and from every other number in the set.

The Variance of Age attribute is,

$$\text{Var Age} = 54.571881$$

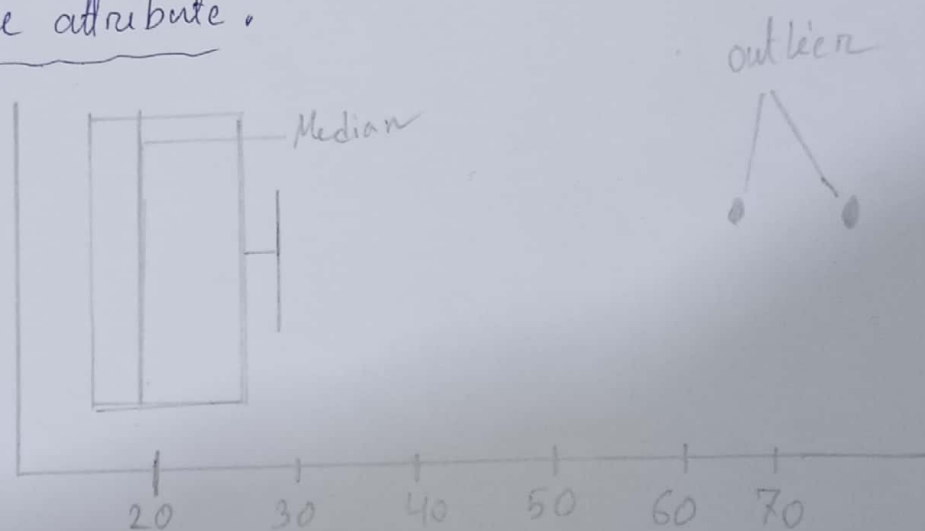
The variance of CGPA attribute is.

$$\text{Var CGPA} = 0.285463$$

▣ Detect Outlier: The outliers are the extreme values within the dataset. That means the outlier data points vary greatly from the expected values — either being much larger or significantly smaller.

For Detecting outlier we use the Box plot method.

- For Age attribute:



- For CGPA

