Experiment 3.3

Student Name: Kanishk Soni **UID: 20BCS9398**
Branch: BE-CSE **Section/Group: 20BCS-DM_705B**
Semester: 6th **Subject Name: Data Mining Lab**
Subject Code: 20CSP-376

1. Aim:

To study Outlier detection using R programming.

2. Objective:

- To create a curve based on prediction using the regression model.

3. Code and Output:

PROGRAM

```
#creating the data containing 500 random values

data <– rnorm(500)


#adding 10 random outliers to this data.

data[1:10] <– c(46,9,15,–90,42,50,–82,74,61,–32)


#draw boxpolot and an outlier is defined as a data point that is located outside the whiskers of the box plot.

boxplot(data)


#remove the outlier of the provided data boxplot.stats() function in R

data <– data[!data %in% boxplot.stats(data)$out]


#draw boxplot to verify whether ouliers removed or not
```

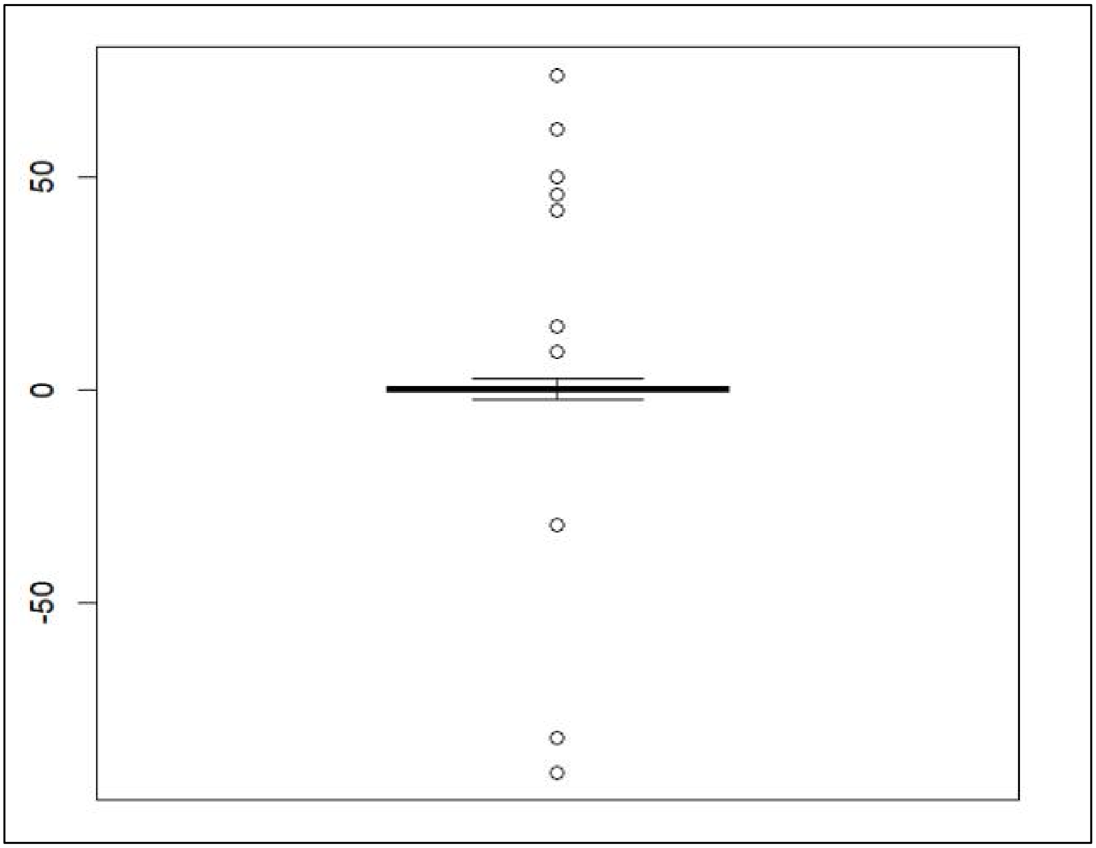
```
boxplot(data)
```

CONSOLE

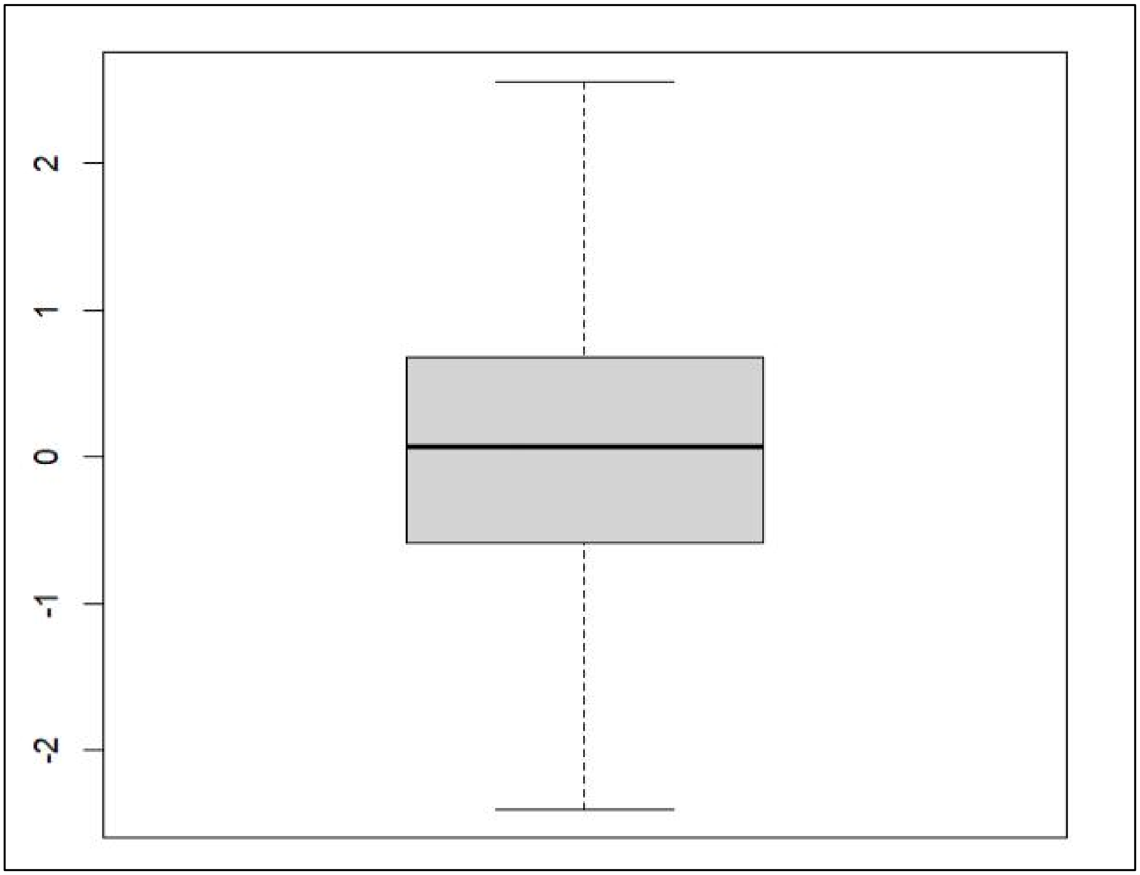
```
> data <– rnorm(500)
> #adding 10 random outliers to this data.
> data[1:10] <– c(46,9,15,–90,42,50,–82,74,61,–32)
> #draw boxpolot and an outlier is defined as a data point that is located outside the whiskers of the box plot.
> boxplot(data)
> #remove the outlier of the provided data boxplot.stats() function in R
> data <– data[!data %in% boxplot.stats(data)$out]
> #draw boxplot to verify whether ouliers removed or not
> boxplot(data)
```

4. Output:

Detecting outliers outside the main dataset:



Dealing with the outliers and after removing:



Learning Outcomes:

- Data points far from the dataset's other points are considered outliers. This refers to the data values dispersed among other data values and upsetting the dataset's general distribution.
- Effects of an outlier on a model may result into the model's accuracy being biased, modifies the mean, variance, and other statistical characteristics of the data's overall distribution also the format of the data appears to be skewed.