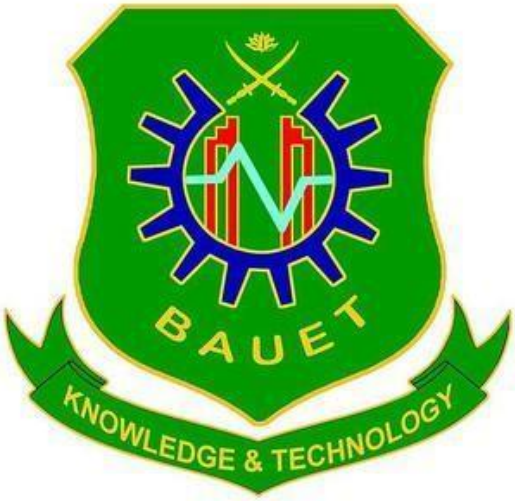
Bangladesh Army University of Engineering & Technology (BAUET)
Qadirabad Cantonment, Natore-6431



Department of Computer Science and Engineering (CSE)

Thesis Title: Outlier detection on data mining

Submitted By:	Submitted to:
Name: MD. Touhid Iqbal Sagar Roll: 18204023 Registration No: 180204023 B.Sc. Engineering Session: 2018-19	Md. Muktar Hossain Sr. Lecturer Dept. of Computer Science & Engineering BAUET

1. Abstract:

In data mining, existing researches mostly based on finding patterns in large datasets and further using it for organizational decision making. However, outlier detection has not yet received as much attention in the data mining field as some other topics like association rules, classification and clustering. So, this paper describes Using scatter plots, Box plot, Z score and IQR interquantile range for outlier detection in data mining.

2. Field of project:

Outlier detection is the process of detecting and subsequently excluding outliers from a given set of data. An outlier may be defined as a piece of data or observation that deviates drastically from the given norm or average of the data set. An outlier may be caused simply by chance, but it may also indicate measurement error or that the given data set has a heavy-tailed distribution.

3. Existing Works:

I have seen some existing works by the help of google scholar, Google, Wikipedia about outlier detection. Many papers have so many algorithms for object detection. I studied many of them and decided to research more about object detection.

4. Proposed Possible Method:

There are various way of outlier detection. I have decided of some techniques for the paper. They are:

I. Scatter plots
II. Box plot
III. Z score
IV. IQR interquantile range.

5. Conclusion:

This paper is based for outlier detection. It helps in Fraud Detection, Text data Detection, Speech Recognition, Traffic Monitoring. Which all are much useful in this modern world.

6. References:

Yu, D., Sheikholeslami, G. and Zang, “A find out: finding outliers in very large datasets”, In Knowledge and Information Systems, 2002, pp.387 - 412.

M.F. Jiang, S.s. Tseng, C. M. Su.,” Two-phase clustering process for outlier detection. pattern recognition letters”, 2001, vol. 22(6-7), pp. 691–700.