



# MINING OF OUTLIER DETECTION IN LARGE CATEGORICAL DATASETS

Mrs. Ramalan Kani K<sup>1</sup>, Ms. N.Radhika<sup>2</sup>

<sup>1</sup>M.TECH Student, Department of computer Science and Engineering, PRIST University, Trichy

<sup>2</sup>Asst.Professor, Department of Computer Science and Engineering, PRIST University, Trichy

---

## Abstract

Outlier detection will typically be thought of as a pre-processing step for locating, throughout a data set, those objects that do not fits well-defined notions of expected behaviour. it is vital in process for locating novel or isolated events, anomalies, vicious actions, exceptional phenomena, etc. We have got an inclination to reinvestigating outlier detection for categorical data sets. This draw back is very hard owing to the matter of shaping a pregnant similarity live for categorical data. Throughout this paper, we have got an inclination to propose an accurate definition of outliers associated academic degree improvement model of outlier detection, via a replacement construct of holoentropy that takes each entropy and total correlation into thought. Supported this model, we have got an inclination to stipulate a perform for the outlier issue of associate object that's fully determined by the item itself and will be updated expeditiously. We have got AN inclination to propose 2 wise 1-parameter outlier detection ways in which, named ITB-SS and ITB-SP, that need no user-defined parameters for deciding whether or not or not or not associate object is associate outlier. Users would like solely supply the number of outliers they need to notice. Experimental results show that ITB-SS and ITB-SP square measure easier and economical than thought ways in which and will be accustomed agitate each vast and high-dimensional data sets wherever existing algorithms fail. Other ways like possibility, Hyper graph theory or agglomeration ways in which goes fail in outlier detection in categorical data. We have got an inclination to area unit measure the outlier detection pattern entropy and total correlation.

**Keywords:** Outlier detection, holoentropy, total correlation, outlier factor, attributes weighting

---

Full Text: [www.ijcsma.com/publications/march2014/V2I310.pdf](http://www.ijcsma.com/publications/march2014/V2I310.pdf)