AENSI Journals

Advances in Environmental Biology

ISSN-1995-0756 EISSN-1998-1066

Journal home page: http://www.aensiweb.com/aeb.html

Using Outlier Detection for Classification of Analysts' Equity

[1]Mohammad Ali Aghaei, [2]Zinat Ansari, [3]Saber Maleki

[1]Professor in Department of Accounting, Science & Research Branch of Bushehr, Islamic Azad University, Bushehr, Iran
[2]Department of Accounting, Shiraz Branch, Islamic Azad University, Shiraz, Iran
[3]Department of Accounting, Science & Research Branch of Bushehr, Islamic Azad University, Bushehr, Iran

ARTICLE INFO

Article history:
Received 15 April 2014
Received in revised form 22 May 2014
Accepted 25 May 2014
Available online 15 June 2014

Key words:
outlier detection, nonlinear projections, genetic algorithms

ABSTRACT

Outlier detection is a primary step in many data-mining applications. In presence of outliers, special attention should be taken to assure the robustness of the used estimators. Outlier detection for Data Mining is often based on distance measures, clustering and spatial methods. In the paper we used 11 input that involve Cash, Short-Term Investments, Notes Receivable, Inventory, Spare Parts, Inventory Stock And Other Inventory, Advance Payment , Long-Term Assets, Notes Payable, Prepaid, Long-Term Liability that applied for classification of equity by outlier detection.

To Cite This Article: Mohammad Ali Aghaei, Zinat Ansari, Saber Maleki, Using Outlier Detection for Classification of Analysts' Equity. ***Adv. Environ. Biol., 8(12),*** *41-43, 2014*

INTRODUCTION

Data mining is the non-trivial method of identifying valid, novel, potentially useful, and finally understandable patterns in data [1]. Now, data mining is becoming an important tool to convert the data into information. It is commonly used in a wide series of profiling practices, such as marketing, fraud detection and scientific discovery [2]. Data mining is the method of extracting patterns from data. It can be used to uncover patterns in data but is often carried out only on sample of data [3]. The mining process will be ineffective if the samples are not good representation of the larger body of the data. The discovery of a particular pattern in a particular set of data does not necessarily mean that pattern is found elsewhere in the larger data from which that sample was drawn. An important part of the method is the verification and validation of patterns on other samples of data. A primary reason for using data mining is to assist in the analysis of collection of observations of behavior [4]. Cluster analysis or clustering is the assignment of a set of observations into subsets called clusters so that observations in the same clusters are similar in some sense. It is a useful technique for the discovery of data distribution and patterns in the original data. The goal of clustering technique is to find out both the dense and the sparse region in a data set. It is a method of unsupervised learning and a common technique for statistical data analysis used in many fields, including machine learning, data mining, pattern recognition, image analysis and bioinformatics [5].

It is an important technique used for outlier analysis. Outlier detection based on clustering approach provides new positive results. Clustering algorithms are used for outlier detection, where outliers (values that are "far away" from any cluster) may be more interesting than common cases. Clustering is a challenging field of research in which its potential applications pose their own requirements. Outliers detection is an outstanding data mining task, referred to as outlier mining. Outliers are objects that do not comply with the general behavior of the data. By definition, outliers are rare occurrences and hence represent a small portion of the data. Outlier detection has direct applications in a wide variety of domains such as mining for anomalies to detect network intrusions, fraud detection in mobile phone industry and recently for detecting terrorism related activities [6].

In the paper we used 11 input that involve Cash ،Short-Term Investments ،Notes Receivable, Inventory، Spare Parts, Inventory Stock And Other Inventory, Advance Payment , Long-Term Assets, Notes Payable, Prepaid , Long-Term Liability that applied for clustering of equity by outlier detection.

Corresponding Author: Saber Maleki, Department of Accounting, Science & Research Branch of Bushehr, Islamic Azad University, Bushehr, Iran,
Tel: +989144190785, E-mail: s.maleki@mehrpetrochem.com

MATERIALS AND METHOD

Material:

In the study area used 12 charactristics that is following:

Table 1: Input data.

Elements	Notes receivable	Inventory	Inventory stock and other inventory	Advance payment	Long-term assets	Notes payable	prepaid	spare parts	Cash	Long-term liability	Short-term Investments	equity
Maximum	5253206	688701	2542277	2521124	17363330	5695291	3139402	9726510	1182705	3001470	3564611	5253206
Minimum	3885	0	0	39	54030	16278	3068	0	900	0	-3868050	3885
Average	1261087	120072	636372	326940	4689123	1478423	704671	1653372	214590	368339	270927	1261087
STDEV	1399529	163515	623930	655760	4879249	1763770	791798	2447987	308493	759892	1794244	1399529

An outlier is an observation that appears to deviate markedly from other observations in the sample. Identification of potential outliers is important for the following reasons.

1. An outlier may indicate bad data. For example, the data may have been coded incorrectly or an experiment may not have been run correctly. If it can be determined that an outlying point is in fact erroneous, then the outlying value should be deleted from the analysis (or corrected if possible).

2. In some cases, it may not be possible to determine if an outlying point is bad data. Outliers may be due to random variation or may indicate something scientifically interesting. In any event, we typically do not want to simply delete the outlying observation. However, if the data contains significant outliers, we may need to consider the use of robust statistical techniques.

RESULTS AND DISCUSSION

The results of the recearch is show in Figure 1.

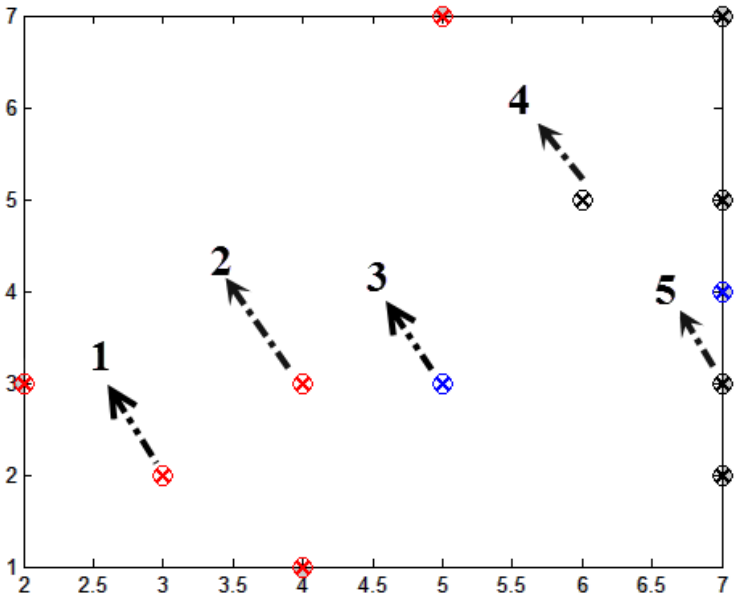

Fig. 1: Classification of equity by outlier detection.

Conclusion:

In the research were used 11 input that involve Cash ،Short-Term Investments ،Notes Receivable, Inventory, Spare Parts, Inventory Stock And Other Inventory, Advance Payment, Long-Term Assets, Notes Payable, Prepaid, Long-Term Liability that applied for classification of equity. For classification of data used outlier detection method. According to results, there are 5 classes for the data.

REFERENCES

[1] Pujari, A.K., 2001. Data Mining Techniques, Universities Press (India) Private Limited.

[2] Osouli, N., E. Khoshbakht, Z. Ansari, M. Kazemi, 2014. Using Self-Organizing Map (SOM) algorithm for Analysts' equities clustering, Advances in Environmental Biology, 8(1): 185-189.
[3] Momeni, A., S. Maleki, R. Khajeh, 2014. Analysts' Equity Forecasts Using of Multi-layer Perception (MLP), Advances in Environmental Biology, 8(1): 207-211.
[4] Osouli, N., Z. Ansari, M. Kazemi, 2014. Using Radial Basis Function (RBF) for Analysts' equities forecasts, Advances in Environmental Biology, 8(1): 217-220.
[5] Esfahaniyan, M., M. Aghabarari, E. Safari, 2013. A Survey of the Relationship between Earnings Management and the Cost of Capital in Companies Listed on the Tehran Stock Exchange, Advances in Environmental Biology, 7(10): 2775-2781.
[6] Challagalla, A., S.S. Shivaji Dhiraj, D.V.L.N Somayajulu, T. Shaji Mathew, S. Tiwari, Syed Sharique A. "Privacy Preserving Outlier Detection Using Hierarchical Clustering Methods, 2010. 34th Annual IEEE Computer Software and Applications Conference Workshops.