International Journal of Business Analytics and Intelligence
7 (1) 2019, 08-23
http://publishingindia.com/ijbai/

Submitted: 12 January, 2019
Accepted: 15 February, 2019

The Use of Data Mining Techniques in Environmental Accounting: A Comparison of Public and Private Sector Companies in India

Hamid Ravan Paknoodezh*

Abstract

Environmental accounting as an essential tool for understanding the role played by the natural environment in the economy is useful for business decisions, especially for proactive environmental management activities. Since recent technological advancement paved the way for the use of big data to assist companies in the decision-making process and one of the best methods to the exploited large dataset is data mining, this study aims to examine the level of data mining techniques in environmental accounting within public and private sector companies in India. This study covering the States of Haryana for North, West Bengal for East, Maharashtra for West and Kerala as representative of South. With the use of structured questionnaires in the soft and hard copy at last 100 managers and 243 accountants randomly were selected. Six hypothesis has been considered that were evaluated in the statements with a five-point Likert scale. In this study, we used the classification and regression techniques as appropriate tools for data mining. The results of this study confirmed the null hypothesis 'there is no significant difference existed in the level of data mining between the public and private sector industry in India' rejects. Therefore, the levels of data mining within environmental accounting in terms of most aspects significantly is higher in the private sector than public sector companies in India.

Keywords: Environmental Accounting, Environmental Management, Data Mining, Industry in India

Introduction

Background of the Study

Today environmental accounting perspective has attracted much attention to reducing or avoiding those costs of the corporate product while at the same time improving environmental quality Yu (2007). Modern accounting is not only concerned with record-keeping and reporting of information to the investors but also aims at fulfilling the information needs of a wide range of internal and external shareholders. It is now considered as a service activity that is estimating and accounting the costs of environmental impacts and rapidly developing an area of management, accounting, and finance. Environmental accounting is a function to provide quantitative information primarily of financial nature, about economic activities that are intended to be useful in making an economic decision Alshhadat (2018). Due to growing public concern about the alarming impact of industrial activities on nature, companies are under force from both the government and the society to reduce adverse impacts of their activities on the environment Chinchuluun (2010). The performance of an institute is now being judged not only by its financial results but also with regard to its contribution to protect and improve the environment Common (1996). In environmental accounting, check the quality of the large size of the estimated data matrix is a difficult task. But the data mining system at the same time is capable of dealing with the temporal and spatial data simultaneously that have unique characteristics for determining. Environmental accounting with the use of proper methodologies like data mining able to correctly measure the environmental impact thus in order to report the environmental cost of the activity of an organization Gilberto (2008). Data mining is defined as the process of sorting through large data sets to recognize patterns and establish relationships to solve problems through data analysis Peter (2013). Data mining tools allow enterprises to predict future trends. In many research areas, data mining techniques are used including mathematics, cybernetics, genetics and

* Department of Commerce, Kerala university, Kerala, India. Email: Hamid.Ravanpakwin7@gmail.com

marketing Shah (2014). While data mining techniques are a means to drive efficiencies and predict customer behaviour, if used correctly, a business can set itself apart from its competition through the use of predictive analysis. In general, the benefits of data mining come from the ability to uncover hidden patterns and relationships in data that can be used to make predictions that impact businesses Rouse (2017). The growth and development of business and increasing of competition are the cause of increasing complexity in business. Also, managers and decision-makers need to predominate on the details of the business, and therefore in this regard data mining became necessary. The criteria for evaluating performance and usefulness of the Environmental Accounting System are the ability of the system to provide sufficient and necessary information from data mining. After all, the objectives of this research assess the level of data mining within an Environmental Accounting System of public and private sector companies in India.

Literature Review

Over the past years, there has been a dramatic increase in the number of studies which have focused on data mining in environmental accounting. Seifert (2004) examined that data mining is a useful tool at the disposal of all companies. Data mining involves the use of sophisticated analytical tools to discover valid patterns and relationships within the massive dataset Seifert (2004). In other words, data mining is an operation beyond the collection and management of data. This operation also involves analysis as well as forecasting. Pechenizkiy (2005) have argued that some of the experts believe that concepts of data mining techniques are a stage of the comprehensive process, which is called knowledge discovery within the database. The other steps in this extensive process involve data cleaning, data integration, data selection, transformation, and pattern evaluation. Data mining technology makes the connection between many technical areas including databases statistics, machine learning, and human-computer interface Pechenizkiy (2010). In the business world, financial data are considered as a strategic asset. Financial data are collected and stored by institutions such as banks, stock exchange, tax agencies, specialized databases related to auditors, accountants, etc. Data mining methods in financial information can solve the problems of classification and prediction used to facilitate the decision-making process. The importance of data mining in finance and accounting can have a wide range that includes fraud detection in increasing the profitability of the business unit. Kirkos (2007) mentioned that many professional organizations recognize the importance of data mining. American Institute of Certified Public Accountants (AICPA), introduced data mining as one of the top ten technologies of the future. The American Association of Internal Auditors also included data mining in the list of their research priorities. The possible application of data mining within environmental accounting information leads to the identification of the level of utilization of data mining in market segmentation, customer churn, fraud detection, market basket analysis, interactive marketing and trend analysis Kirkos (2007). In terms of data mining application areas refer to surveys such as those conducted by Sirikulvadhana in (2002) have shown that classification and prediction capabilities of data mining techniques are used to facilitate the audit process, also help assess the credit risk, and to predict the performance of the company. Auditing is known as an old profession that has realized, in recent years, sharp increases in trading volume and complexity of financial and non-financial data to their client. To audit these companies, auditors are dealing with bulk data and complex structured data. As a result, auditors cannot only rely on tools for reporting the audit process Sirikulvadhana (2002). Data mining can be a useful tool to automatically extract the information from large volumes of data, although an adaptation of data mining in the audit process is a relatively new technique. However, data mining has shown, in many commercial applications related to the audit such as fraud detection and forensic accounting, to have a great advantage. The tasks of data mining have been widely investigated by Sumathi (2006) that noted data mining is in the early stages of its life. A large part of industries including the financial, healthcare, production, transportation industry are using data mining tools and techniques to achieve their goals. For this purpose, they are using old data. They are taking advantage of shifting techniques and technology, mathematics and statistical analysis of information available in the database. With the help of data mining, the analyzers can detect critical events, business communication processes, and abnormalities that need not be considered Sivanandam (2006). In

business, data mining is used to discover the relationships between the data and also used to make better decisions. Data mining also helps to determine the spot sales trend, more efficient marketing and help in predicting customer loyalty. In the context of data mining in the environmental accounting information, Thuraisingham (2000) pointed out that data mining within environmental accounting information is particularly the process of collecting and analyzing environmental accounting data, and presenting them in a way that generates environmental information and knowledge. The knowledge gained through the analysis of this information can be used to further goals of decision-making within organizations. Data mining activities include summarizing, comparing, analyzing, forecasting and estimating Thuraisingham (2000). Data mining enables organizations to apply updated statistical tools and software for analyzing the environmental data as well as using knowledge management by database management for the extraction of information from large databases. Mendes (2017) observed that the adoption of this technology increases the responsibility of accountants and auditors. With the help of this technology, environmental accounting system will be able to produce complete, timely and related environmental information required for decision-making. The users of accounting information mainly need current and ongoing information Mendes (2017). Crespo (2005) investigated the aspect that data mining is not just a tool to keep track of transactions, but it can also complete testing of the system, as well as undertaking the necessary controls to ensure that the organization can produce accurate financial statements according to environmental information Crespo (2005). The main survey in an issue like A. Amani (2017) studies examined that data mining increases the ability of environmental accounting information to take an effective role in the collection of transaction data, providing information to decision-makers, and collaborate on internal controls. Environmental accounting information maintains volumes of transactions, and it's suggested to use them as primary sources of environmental information to carry out organizational purposes Amani (2017).

Therefore, based on the theoretical basis of our study hypothesis is presented as follows:

H1: The environmental marketing helps to products that are presumed to be environmentally safe.
H2: Environmental accounting predicts which customers are likely to leave your company and go to your competitor.
H3: The environmental accounting system helps to identify the manipulation of financial data, intentional deception, and misappropriation of company natural assets.

H4: Environmental accounting helps to find out which services were accepted and used more by the customer to reduce environmental pollution.

H5: Environmental accounting system provides a basis for interactive marketing in the business.
H6: Environmental accounting system helps to analyses the business trends, according to environmental considerations in the course of time.

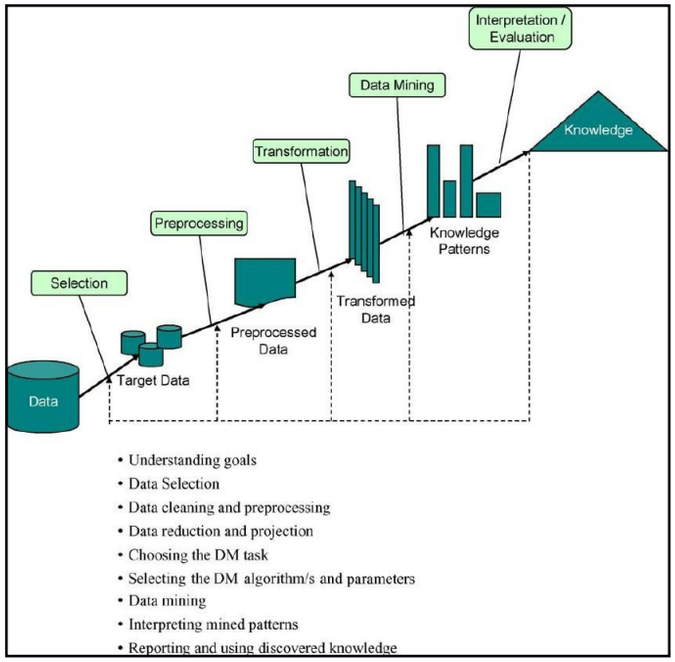
Research Methodology

Sampling Procedure and Data Collection

The study was conducted in India, covering the States of Haryana for North, West Bengal for East, Maharashtra for West and Kerala as representative of South. In this study two separate sets of structured questionnaires were administered, one for managers and one for accountants in the form of soft and hard copy. The sample of study consists of 100 managers, and 243 accountants were selected at random from the public sector (BPCL) and private (AOL) sector companies.

Research Design

Environmental accounting information, manages and exchanges the generated reports and provides services to different systems of the organization. These include the environmental systems, financial systems, management accounting and payroll systems as well as branches of accounting systems. As the model shows, this computerized information system is used as a data storage centre, which contains information about different parts of the accounting system. Using data mining technology involves a combination of information and communications technology (ICT), statistical data analysis tools and knowledge management, which increases the knowledge of decision-makers about accounting A.Rahman (2008).



Source: Utilization of data mining technology within the environmental accounting information in the public sector, Rahman (2008)

Fig. 1: Data Mining used within the Environmental Accounting Information

The goal of data mining is to analyse massive amounts of data to look for patterns that can help predict future results. It can be used to predict customer behaviour, market oscillations, find the best business moves, and so on. The procedure of data mining involves:

Data Mining Process

- *Identification of Objective:* At this stage, it will be determined what the user needs and the level of information he intends to obtain from the database. Also, it is required to understand research objectives clearly and find out what are the needs, on the other hand, the study what want to achieve with these data.
- *Selection of Data:* The data must be based on criteria of objective selection. This involves going over the data to assess its quality and prepare what you need for the following phases. Assess finding the resources, assumptions, constraints and other important factors which should be considered.
- *Data Preparation:* The function of this stage of the process will be the preparation of data and identification of redundant variables used.
- *Evaluation of Data*: The general framework of this stage is the criteria such as the type of data distribution, characteristics and structure of the database and the general condition of the data etc.
- *The Answer Formwork:* The output of this section is to provide formats in the form of images, graphs and neural networks and so on. Then, using mathematics to identify patterns and use them to build different models.
- *Selection Tools:* At this point, appropriate tools for data mining are selected. Data mining is done by techniques like classification, clustering, regression, association rules, outer detection, sequential patterns, and prediction. In this research, we used the classification and regression techniques. Classification analysis is used to retrieve important and relevant information about data, and metadata. This data mining method helps to classify data in different classes. Regression analysis is the data mining method of identifying and analyzing the relationship between variables. It is used to identify the likelihood of a specific variable, given the presence of other variables.
- *Modelling:* Data mining process begins at this stage, which includes search patterns in data sets, classification, and evaluation of the data.
- *Validated Results:* This stage includes the test patterns.
- *Presenting the Results*: The results of this section are the final report to the users.
- *Use the Results:* The purpose of data mining is using the obtained results for decision-making, policy-making, and forecasting to create a better and new situation (Data Mining Tutorial: Process, Techniques, Tools & Examples, 2018; Scott, 2018; Ruxandra, 2013; W. Seifert, 2004).

Results

In this research, we focus on the aspects that are most important components to assess the level of data mining within an Environmental Accounting System in Indian public and private sector companies. In this respect, six

hypotheses has been considered that were evaluated in the statements with a five-point Likert scale, from strongly disagree to agree strongly. The statements measured in classification method are the opinion of managers and accountants regarding the level of data mining within an Environmental Accounting System in Indian companies.

Table 1 shows the number and percentage of public and private sector managers having an opinion regarding the level of data mining in Indian companies. From the table 1, for example, it can be seen that the six aspects of data mining received zero answers for 'strongly disagree' and 'disagree' in both public as well as private sector. In the case of market segmentation, forty per cent of responses received from the public and thirty present from the private sector was 'neutral, ' and sixty per cent of public respondents and seventy per cent of private sector respondents were either 'agree,' or 'strongly agree.'

Table 1: Distribution of Managers of Public and Private Sector in India by Their Responses to the Level of Existence of Various Aspects of Data Mining in Their Company

Aspects of Data Mining	*Response*	*Public*		*Private*	
		N	*%*	*N*	*%*
Market segmentation	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	20	40.00	15	30.00
	Agree	25	50.00	30	60.00
	Strongly Agree	5	10.00	5	10.00
Customer churn	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	25	50.00	20	40.00
	Agree	20	40.00	25	50.00
	Strongly Agree	5	10.00	5	10.00
Fraud detection	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	15	30.00	5	10.00
	Agree	25	50.00	35	70.00
	Strongly Agree	10	20.00	10	20.00
Market basket analysis	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	20	40.00	10	20.00
	Agree	30	60.00	25	50.00
	Strongly Agree	0	0.00	15	30.00
Interactive marketing	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	20	40.00	20	40.00
	Agree	30	60.00	30	60.00
	Strongly Agree	0	0.00	0	0.00
Trend analysis	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	10	20.00	5	10.00
	Agree	35	70.00	20	40.00
	Strongly Agree	5	10.00	25	50.00

Source: primary (2017)

Also, the responses in regard interactive marketing are as follows; forty per cent of both sectors respondents expressed 'neutral' response, sixty per cent of responses in both sectors is showing 'agree' and none of them responded 'strongly agree' in both sectors. The last aspect measured is the trend analysis, which respectively received twenty per cent and ten per cent 'neutral' response from public and private sector companies. Eighty per cent of the public sector and ninety per cent of private sector respondents were either 'agree,' or 'strongly agree.'

Table 2: Mean Opinion Scores of Managers About the Level of Various Aspects of Data Mining in Companies with One Sample T-Test And Two Independent Samples T-test

Aspects of data mining	*Public*		*Private*		*t*	*Sig*
	Mean	*SD*	*Mean*	*SD*		
Market segmentation	3.70*	0.65	3.80*	0.61	-0.798	0.427
Customer churn	3.60*	0.67	3.70*	0.65	-0.759	0.450
Fraud detection	3.90*	0.71	4.10*	0.54	-1.585	0.116
Market basket analysis	3.60*	0.50	4.10*	0.71	-4.096	0.000
Interactive marketing	3.60*	0.50	3.60*	0.50	0.000	1.000
Trend analysis	3.90*	0.54	4.40*	0.67	-4.096	0.000

* Significantly greater than 3.00 as per one sample t-test

Source: primary data (2017)

Table 2 shows the mean opinion scores of managers regarding the level of data mining within the environmental accounting system (EAS) in public and private sector companies in India separately. It can be found that all mean scores are greater than three as the significance levels of one sample t-test with test value 3.00 are smaller than 0.05. It means that according to managers of both public and private sector companies the level of data mining is above average. From the two independent sample t-test, it can be understood that managers have a difference of opinion in two attributes, namely market basket analysis and trend analysis as the significance levels are lesser than 0.05. From the comparison of mean scores of the public and private sector, it can be seen that the market basket analysis and trend analysis are better in the private sector as the mean scores are 4.10 and 4.40 respectively.

Table 3: Distribution of Accountant of Public and Private Sector in India by Their Responses to the Level of Existence of Various Aspects of Data Mining in Their Company

Aspects of Data Mining	*Response*	*Public*		*Private*	
		N	*%*	*N*	*%*
Market segmentation	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	29	23.58	36	30.00
	Agree	94	76.42	60	50.00
	Strongly Agree	0	0.00	24	20.00
Customer churn	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	62	50.41	60	50.00
	Agree	40	32.52	36	30.00
	Strongly Agree	21	17.07	24	20.00
Fraud detection	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	10	8.13	0	0.00
	Agree	83	67.48	60	50.00
	Strongly Agree	30	24.39	60	50.00
Market basket analysis	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	22	17.89	12	10.00
	Agree	101	82.11	96	80.00
	Strongly Agree	0	0.00	12	10.00
Interactive marketing	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	21	17.07	24	20.00
	Agree	93	75.61	72	60.00
	Strongly Agree	9	7.32	24	20.00
Trend analysis	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	11	8.94	0	0.00
	Agree	72	58.54	60	50.00
	Strongly Agree	40	32.52	60	50.00

Source: primary data (2017)

Table 3 shows the number and percentage of public and private sector accountants by their opinion regarding the level of data mining in India. From the table, it can be seen that the six aspects of data mining received zero answers for 'strongly disagree' and 'disagree' in both public as well as private sector. In the case of market segmentation 23.58 per cent of responses received from public and 30.00 per cent from the private sector was 'neutral,' and 76.42 per cent of public respondents were 'agree' and no responses received regarding 'strongly agree' and 70.00 per cent of private-sector respondents were either 'agree', or 'strongly agree'. For the aspect customer churn, the percentage of neutral responses received from public and private sector accountants are 50.41 and 50.00 respectively. It can also be seen that 49.59 per cent of public sector accountants and 50.00 per cent of private-sector accountants were either 'agree,' or 'strongly agree.' The third aspect is fraud detection, which has received 8.13 per cent of neutral responses from public sector accountants and 0.00 per cent of private-sector accountants. At the same time, 91.87 per cent of the public sector and 100.00 per cent of private-sector accountants respectively were either 'agree,' or 'strongly agree. The last aspect measured is the trend analysis, which received 8.94 per cent and 0.00 per cent neutral response respectively from public and private sector companies. It can also be seen that 91.06 per cent of the public sector and 100.00 per cent of private-sector respondents were either 'agree,' or 'strongly agree.' The responses of accountants were converted into scores and are presented in Table 4.

Table 4: Mean Opinion Scores of Accountant Regarding the Level of Existence of Various Aspects of Data Mining within EAS in Public and Private Companies with a Two-Sample T-test

Accountant	*Public*		*Private*		*t*	*Sig*
	Mean	*SD*	*Mean*	*SD*		
Market segmentation	3.76*	0.43	3.90*	0.70	-1.826	0.069
Customer churn	3.67*	0.75	3.70*	0.78	-0.338	0.736
Fraud detection	4.16*	0.55	4.50*	0.50	-4.997	0.000
Market basket analysis	3.82*	0.39	4.00*	0.45	3.337	0.001
Interactive marketing	3.90*	0.49	4.00*	0.64	1.347	0.179
Trend analysis	4.24*	0.60	4.50*	0.50	-3.712	0.000

* significantly greater than 3.00 as per one sample t-test

Source: primary data (2017)

Table 4 shows the mean opinion scores of accountants regarding the level of data mining within the Environmental Accounting System (EAS) in public and private sector companies in India separately. It can be found that all mean scores are greater than three as the significance levels of one sample t-test with test value 3.00 are smaller than 0.05. It means that according to accountants of both public and private sector companies the level of data mining is above average. From the two independent sample t-test, it can be understood that accountants have a difference of opinion in three attributes, namely fraud detection, market basket analysis, and trend analysis as the significance levels are lesser than 0.05. From the comparison of mean scores of the public and private sector, it can be seen that the fraud detection, market basket analysis, and trend analysis are better in the private sector as the mean scores 4.50, 4.00 and 4.50 respectively which is higher than that of public sector companies.

Table 5: Distribution of Total Respondent (both managers and accountants) of Public and Private Sector in India by their Responses to the Level of Existence of Various Aspects of Data Mining in Their Company

Aspects of Data Mining	*Response*	*Public*		*Private*	
		N	*%*	*N*	*%*
Market segmentation	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	49	28.32	51	30.00
	Agree	119	68.79	90	52.94
	Strongly Agree	5	2.89	29	17.06
Customer churn	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	87	50.29	80	47.06
	Agree	60	34.68	61	35.88
	Strongly Agree	26	15.03	29	17.06
Fraud detection	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	25	14.45	5	2.94
	Agree	108	62.43	95	55.88
	Strongly Agree	40	23.12	70	41.18
Market basket analysis	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	42	24.28	22	12.94
	Agree	131	75.72	121	71.18
	Strongly Agree	0	0.00	27	15.88
Interactive marketing	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	41	23.70	44	25.88
	Agree	123	71.10	102	60.00
	Strongly Agree	9	5.20	24	14.12
Trend analysis	Strongly Disagree	0	0.00	0	0.00
	Disagree	0	0.00	0	0.00
	Neutral	21	12.14	5	2.94
	Agree	107	61.85	80	47.06
	Strongly Agree	45	26.01	85	50.00

Source: primary data (2017)

Table 5 shows the number and percentage of public and private sector respondents (both managers and accountants) by their opinion regarding the level of data mining in Indian companies. From the table 5, it can be seen that in the case of market segmentation, 28.32 per cent of responses received from the public and 30.00 per cent from the private sector was 'neutral', and 71.68 per cent of public and 70.00 per cent of private sector respondents were either 'agree,' or 'strongly agree.' And also, the aspect of fraud detection has received 14.45 per cent and 2.94 per cent 'neutral' opinion from the public sector and private sector respondents respectively. The table also indicated that 85.55 per cent of public and 97.06 per cent of private-sector respondents were either 'agree,' or 'strongly agree.' The responses regarding the aspect market basket analysis are as follows; 24.28 per cent of public and 12.94 per cent of private-sector respondents expressed 'neutral' opinion. The responses received have shown 75.72 per cent of the public sector, and 87.06 per cent of private-sector respondents were either 'agree,' or 'strongly agree.' The last aspect measured is the trend analysis, which respectively received 12.14 per cent and 2.94 per cent 'neutral' response from the respondents of public and private sector companies. The responses received have shown 87.86 per cent of the public sector, and 97.06 per cent of private sector respondents were either 'agree,' or 'strongly agree.' The responses of accountants were converted into scores and are presented in Table 6.

Table 6: Mean Opinion Scores of the Total Respondent (both managers and accountants) About the Level of Existence of Various Aspects of the Level of Data Mining within EAS in Public and Private Companies with Two Sample T-test

Combined	*Public*		*Private*		*t*	*Sig*
	Mean	*SD*	*Mean*	*SD*		
Market segmentation	3.75*	0.50	3.87*	0.68	-1.950	0.052
Customer churn	3.65*	0.73	3.70*	0.75	-0.661	0.509
Fraud detection	4.09*	0.61	4.38*	0.55	-4.738	0.000
Market basket analysis	3.76*	0.43	4.03*	0.54	-5.182	0.000
Interactive marketing	3.82*	0.51	3.88*	0.62	-1.099	0.273
Trend analysis	4.14*	0.60	4.47*	0.56	-5.291	0.000

* Significantly greater than 3.00 as per two-sample t-test

Source: primary data (2017)

Table 6 shows the opinion scores of both managers and accountants regarding the level of data mining within the Environmental Accounting System (EAS) in public and private sector companies in India. It can be found that all mean scores are greater than three as the significance levels of two sample t-test with test value 3.00 are smaller than 0.05. It means that according to respondents of both public and private sector companies the level of data mining is above average in their company. From the two independent sample t-test, it can be understood that respondent has a difference of opinion in three attributes, namely fraud detection, market basket analysis, and trend analysis as the significance levels are lesser than 0.05. From the comparison of mean scores of the public and private sector, it can be seen that the fraud detection, market basket analysis, and trend analysis are better in the private sector as the mean scores are 4.38, 4.03 and 4.47 respectively which greater than that public-sector companies. To find the level of data mining within the Environmental Accounting System in India, the elements of data mining were combined by taking a weighted average using Principal Components Analysis (PCA). Factor loadings were taken as weights for each element of the data mining. Factor loading was calculated from the responses collected from the managers of public and private sector companies. Table 7 presents the weights of elements of the data mining.

Table 7: Weight for the Elements of Data Mining

No.	*Elements of Data Mining*	*Weight*
1	Market segmentation	0.596
2	Customer churn	0.559
3	Fraud detection	0.323
4	Market basket analysis	0.805
5	Interactive marketing	0.743
6	Trend analysis	0.600

Source: primary data (2017)

From Table 7 it can be seen that 'market segmentation' with the weight of 0.805 is the most important aspect of data mining within the Environmental Accounting System in India. 'Interactive marketing' with a weighted score of 0.743 is the second important aspect and the 'trend analysis' has got the third rank with the weight score of 0.600. The findings of this table indicate that, as per the view of respondents, the elements which obtained higher

weight score, will have a higher manifestation in the level of data mining within the Environmental Accounting System than other elements with lower factor loadings and vice versa. Using the above weights, the opinion scores of respondents about the level of data mining within EAS was computed, and descriptive statistics of the estimated score are presented in Table 8.

Table 8: Descriptive Statistics of Opinion Score of Respondents about the Level of Data Mining within EAS by the Opinion of Managers and Accountants

Statistic	*Data Mining*	
	Managers	*Accountant*
Mean	3.81*	3.98*
Median	3.86	4.00
Mode	4.00	4.00
Skewness	0.115	0.647
Kurtosis	-0.732	0.526
Maximum	4.80	4.84
Minimum	3.09	3.48
Range	1.71	1.36
SD	0.47	0.32
t	3.850	
Sig.	0.000	

Source: primary data (2017)

From the table 8, it can be seen that the mean scores of the level of data mining are more than three for both managers as well as accountants. The mean score obtained for the managers is 3.81, and that for accountants is 3.98 which is significantly different as per one sample t-test since the significance levels are found to be lesser than 0.05. The result indicates that the level of data mining is above average in both public and private sector companies in India. The two sample t-test shows that the mean opinion scores of accountants are significantly higher than that of the manager as the significance level is lesser than 0.05. From this result, it can be concluded that the accountants have the opinion that the level of data mining is higher within the Environmental Accounting System as compared to that of the manager.

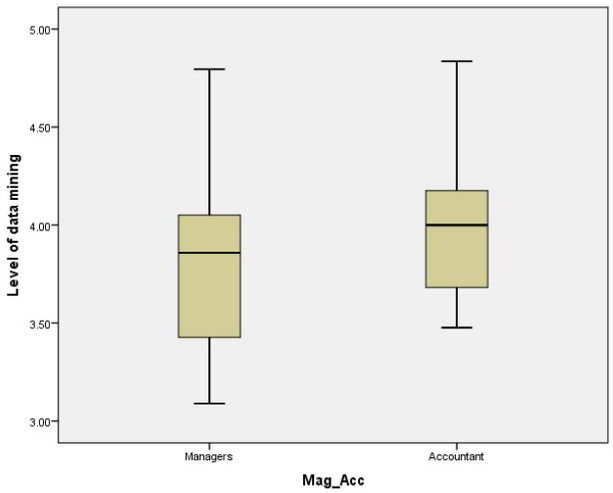


Fig. 2: Box Plot Representing the Mean Score of Elements of the Level of Data Mining by Designation

From Fig. 2 it can be inferred that the level of data mining as per the opinion of managers in companies to an accountant is different. It can be seen that the median of box plot showing that the fifty per cent of responses with positive skewness is acquired the score greater than 3.86, which means that they rely on an environmental accounting system for data mining. The top whisker shows the maximum score of 4.80, and down whisker show, the minimum score of 3.09. Table 8 discerned that the difference between the maximum and a minimum score of manager's opinion is 1.71. The box plot of accountants showing the median score of 3.98, which indicates that managers are less relied on data mining within environmental accounting compared to an accountant. The box plot of accountants showing the positive skewness of responses. The top whisker shows the maximum score of 4.84, and down whisker show, the minimum score of 3.48 and the range difference between a maximum and minimum score of accountant's opinion is 1.36. From the comparison of a median score of managers and accountants, it can be concluded that accountants are more rely on data mining within the environmental accounting system than managers. Also from the comparison of interquartile range (Box-and-Whisker Plot), it can be seen that managers are less consistent than the accountants in relying on data mining within environmental accounting.

Table 9: Descriptive Statistics of the Opinion of Respondents from Public and Private Sector Companies about the Level of Data Mining within EAS by Ownership

Statistics	*DM*	
	Public	*Private*
Mean	3.84	4.03
Median	4.00	4.00
Mode	4.00	4.00
Skewness	-0.552	0.081
Kurtosis	-0.279	-0.122
Maximum	4.41	4.84
Minimum	3.09	3.09
Range	1.32	1.75
SD	0.31	0.41
t	4.670	
Sig.	0.000	

Source: primary data (2017)

Table 9 presents descriptive statistics of the level of data mining within an environmental accounting system as per the opinion of public and private sector companies in India. From the table 9, it can be seen that the mean score obtained for the public sector is 3.84 and that for the private sector is 4.03 which is significantly different from the significant level of t-test is less than 0.05. From this result, it can be concluded that the private sector companies are more relying on an environmental accounting system for data mining.

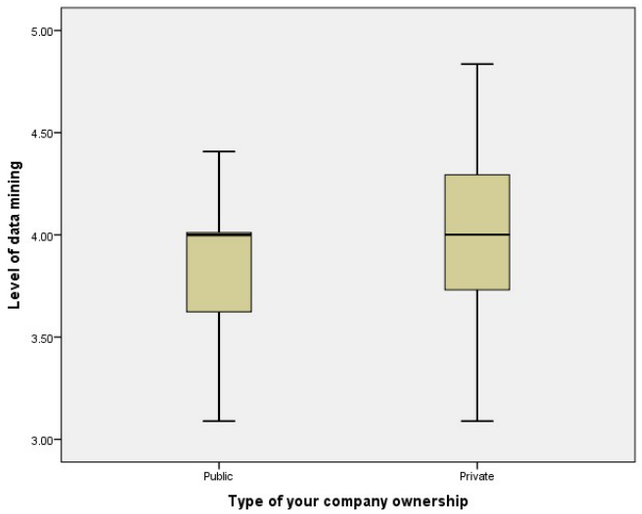


Fig. 3: Box Plot Representing the Mean Score of Elements of the Level of Data Mining by Ownership

From Fig. 3 it can be inferred that the level of data mining as per the opinion of public and private sector companies are different. The box plot of public sector showing the negative skewness of responses regarding the level of data mining within environmental accounting. The median is found to be very close to the upper part of the box indicating that twenty-five per cent of public sector employees has average scores are the median value or little higher than medium. The top whisker shows the maximum score of 4.41, and down whisker show, the minimum score of 3.09. The box showing that fifty per cent of responses is placed between the upper quartile and lower quartile, also the length of the box-and-whisker plot linked to the range of opinion given to the level of data mining within the environmental accounting system. From Table 9 it can be seen that the difference between a maximum and minimum score of respondent's opinion is 1.32. The box plot of private sector companies showing the median score of 4.00, which shows that the fifty per cent of private sector companies is expressing the opinion score greater than this value, which indicates that they strongly rely on data mining within the environmental accounting system. Also, the positive skewness of responses regarding the level of data mining and the top whisker shows the maximum score of 4.84 and down whisker show the minimum score of 3.09. The interquartile range of responses shows that fifty per cent of responses is placed between the upper quartile and lower quartile, also the length of the box-and-whisker plot linked to the range of opinion given to the level of data mining within an environmental accounting system. From Table 9 it can be seen that the range is 1.75, which shows the difference between the maximum and a minimum score of private sector companies' opinion. From the comparison of a median score of public and private sector companies, it can be concluded that private sector companies are more reliant on data mining within environmental accounting. Also from the comparison of interquartile range (Box-and-Whisker Plot), it can be seen that private sector companies are less consistent than the public sector companies in relying on data mining within the environmental accounting system. Table 10 presents the result of the Regression Coefficients of perceptions of respondents about the level of data mining within the environmental accounting system in India. The independent variables selected were gender, age, knowledge, education, ownership, and experience. All independent variables were represented as a dummy variable. So the assumptions of multiple regressions were not verified.

Table 10: Coefficients of the Multiple Regression Model for the Level of Data Mining within EAS

Variables B		*Unstandardized Coefficients*		*Standardized Coefficients*		
		Std. Error	*Beta*	*t*	*Sig.*	
Name	*Dummy variables*					
(Constant)		*4.576*	*0.196*		*23.298*	*0.000*
Designation	manager_d	- 0.091	0.056	-0.11	-1.627	0.105
Type of company ownership	Public_d	-0.09	0.047	-0.12	-1.913	0.057
Gender	male_d	0.093	0.047	0.115	1.99	0.047
Age	Ageless25_d	0.243	0.165	0.239	1.473	0.142
	Age25_d	0.313	0.132	0.388	2.376	0.018
	Age36_d	0.211	0.127	0.239	1.658	0.098
	Age46_d	0.025	0.125	0.028	0.202	0.840
Knowledge about EA	kno_L_d	- 0.691	0.173	-0.324	-3.999	0.000
	kno_A_d	- 0.617	0.115	-0.794	-5.36	0.000
	kno_G_d	- 0.564	0.103	-0.749	-5.468	0.000
Qualification	Edu_prof_d	- 0.546	0.158	-0.538	-3.466	0.001
	Edu_Mas_d	-0.38	0.131	-0.501	-2.892	0.004
	Edu_Bac_d	-0.29	0.131	-0.373	-2.214	0.028
Years of experience	Exp_less5_d	- 0.002	0.099	-0.002	-0.018	0.986
	Exp_5_d	0.098	0.081	0.125	1.207	0.228
	Exp_11_d	0.04	0.08	0.039	0.507	0.612

Source: primary data (2017)

From the table 10, it can be seen that eight dummy variables are found to be significant as their significance levels are less than 0.05. It means that the effects of all eight predictor dummy variables are significant in predicting the perception of respondents on the level of data mining within the environmental accounting. Since out of sixteen dummy variables, fifty per cent of them are significant; this model can be used to predict the level of perception of respondents about the level of data mining.

The validity of this regression model can also be seen from the result of ANOVA given in Table 11 by the fact that the significance level of F-Value is less than 0.05. The R square of the regression model is found to be 0.282. It means that the selected independent variables determine 28.2 per cent variation in the perception of respondents about the level of data mining within the environmental accounting.

Table 11: Result of ANOVA of the Regression Model for Perception of Respondents on the Level of Data Mining within EA

	Sum of Squares	*df*	*Mean Square*	*F*	*Sig.*
Regression	13.616	16	0.851	7.987	0.000
Residual	34.736	326	0.107		
Total	48.352	342			

Source: primary data (2017)

The results of the Regression Coefficients presented in Table 10 provide detailed information about the relative effects of each predictor variable about the respective reference categories (the category with assigned value zero is called the reference category). The variables which have a significant effect include dummy variable representing gender, the respondent 25 to 35 years old, knowledge and qualification in three categories, namely, professionals, masters and bachelors. The table shows that the regression coefficient B obtained is negative which means that the managers' perceptions about the level of data mining within the environmental accounting are lower than that of accountants by an amount of 0.091. As far as ownership is concerned the dummy variable representing public sector (Sig. = 0.057) is not significant in determining the variation in the perceptions of respondents about the level of data mining in concerning the private as the reference category. The respondent with the gender category, the male dummy variable (Sig. = 0.047) is significant in determining the variation in the perceptions of respondents, concerning the female (reference category). However, the table shows that the regression coefficient B obtained the positive perception (B= 0.093) which means that the perception of male respondent about the level of data mining is higher than female respondents by an amount of 0.093. In the case of age, the second age category of respondents is significant 0.018 in determining the variation in the perceptions of respondents is less than 0.05. It can be inferred that the respondent with age between 25 to 35 years received a significant score (B=0.313). Therefore statistically, they have an effect on the level of data mining within environmental accounting. As the regression coefficient B shows the positive value in all categories, it can be said the respondents in all age categories have a high perception of the level of data mining within an environmental accounting compared to the reference category (55+years). It can also be seen from Table 10 that the respondent in all knowledge levels is significant in determining the variation in the perceptions of respondents about the level of data mining. Therefore, statistically, they have an effect on the level of data mining within environmental accounting. As the regression coefficient B shows negative scores in all categories (little knowledge -0.691, average knowledge -0.617, good knowledge -0.564), it can be said that the all respondent categories have a low perception of the level of data mining within an environmental accounting compared to respondents having rich knowledge (reference category). All qualification categories of respondent are significant in determining the variation in the perceptions of respondents about the level of data mining within the environmental accounting. As the regression coefficient B shows negative scores in all categories (professionals -0.546, masters -0.380, bachelors - 0.290), it can be said that the respondents in those categories have a low perception of the level of data mining within an Environmental Accounting compared to less than Bachelor education (reference category). At the same time, respondents' perception of all experience category is not significant as their significance level is higher than 0.05.

Table 12 presents the MCA (Multiple Classification Analysis) adaptations of the perceptions of respondents about the level of data mining within environmental accounting companies derived from the above regression model. The MCA table provides a clear understanding of the effects of predictive characteristics of respondent on the distinction of their perceptions about the level of data mining within an environmental accounting. From Table 12, it has been observed that the unadjusted and adjusted R-values of knowledge about EA (0.335 and 0.553), qualification (0.128 and 0.294), and gender (0.059 and 0.115) differ very much compared to other predictor variables. Unadjusted R values are found to be higher in three variables, namely designation, ownership and experience while adjusted R values are higher for other variables such as gender, age, knowledge, and qualification. It means that the effect of knowledge about environmental accounting, qualification, gender, and age on the perceptions of respondents about the level of data mining within an environmental accounting is augmented by the effect of other variables. At the same time effect of designation, ownership and experience are explained away by the effect of other variables.

Table 12: MCA Table of the Opinion Score about the Level of Data Mining

			Unadjusted		*Adjusted*	
Characteristics		*N*	*Mean Score*	*R*	*Mean Score*	*R*
Designation				0.204		0.110
1	Managers	100	3.815		3.870	

Characteristics		N	Unadjusted Mean Score	Unadjusted R	Adjusted Mean Score	Adjusted R
2	Accountant	243	3.983		3.960	
Type of company ownership				0.245		0.120
1	Public	173	3.843		3.889	
2	Private	170	4.027		3.979	
Gender				0.059		0.115
1	Male	234	3.949		3.964	
2	Female	109	3.902		3.871	
Age				0.266		0.309
1	Less than 25years	56	3.844		3.982	
2	25-35 years	109	4.011		4.052	
3	36-45 years	82	3.954		3.950	
4	46-55 years	75	3.960		3.764	
5	Above 55 years	21	3.607		3.739	
Knowledge about EA				0.335		0.553
1	Little knowledge	11	3.624		3.752	
2	Average knowledge	127	3.885		3.826	
3	Good knowledge	157	3.908		3.879	
4	Rich knowledge	48	4.221		4.443	
Qualification				0.128		0.294
1	Professional quality	56	4.028		3.748	
2	Master's degree	149	3.891		3.915	
3	Bachelor degree	126	3.938		4.004	
4	Others	12	3.979		4.295	
Years of experience				0.273		0.120
1	Less than 5years	82	3.860		3.891	
2	5-10 years	124	3.995		3.991	
3	11-15 years	52	4.102		3.933	
4	Over 15years	85	3.814		3.893	
Full Model				0.531		0.282

Source: primary data (2017)

Since the absolute values of the difference between unadjusted and adjusted R are comparatively lower for age, gender and designation it can be inferred that these variables have an independent effect on the perception of respondents about the level of data mining within environmental accounting of public and private sector in India. Table 12 shows that the *R Square* of the regression model was found to be 0.079_2 which means that 7.95 percent (R Square = R^2 (0.282) *100= 7.95) of variation in the perceptions of respondents about the level of data mining within environmental accounting in the public and private sector was determined by their designation, type of company ownership, gender, age, knowledge about environmental accounting, qualification, and years of experience. From table 12 it is found that the mean opinion score of accountants about the level of data mining is higher in comparison to the managers. The effect of designation on the level of data mining within environmental accounting is found to be dependent on the adjusted R-value 0.110 and unadjusted R- value 0.204. The adjusted R-value is lesser than the unadjusted R-value, these parameters indicating that the effect of the designation on the level of data mining was explained away by the other independent variables. It means that the effect of the designation on the level of data mining within the environmental accounting is not independent. As per the type of company ownership, it can be understood that the mean opinion score of public sector respondents about the level of data mining is lower in comparison to the private sector respondents. The effect of type of ownership at the level of data mining is found to be dependent on other predictor variables as the difference between adjusted R-value 0.120, and unadjusted *R-value* 0.245 is comparatively higher. The adjusted R-value is lesser than the unadjusted R-value, these parameters indicating that the effect of type of ownership on the perception of the level of data mining was explained away by the other independent variables. It means that the effect of type of ownership on the perception of the level of the data mining within the environmental accounting is not independent. As far as the gender of the respondents is concerned, male respondents are more satisfied with the level of data mining than female respondents as their adjusted mean opinion scores are 3.964 and 3.871 respectively. The adjusted R-value is not very much different than the unadjusted R-value which indicates that the effect of

gender on the perception of the respondent about the level of data mining was independent. In the case of age, the respondents with the age group of 25-35 years are more satisfied with the level of data mining companies with an adjusted mean score of 4.052 followed by the youngest age group with an adjusted mean score of 3.982. The opinion scores of respondents in the higher age groups are found to be decreasing steadily. As the difference between adjusted and unadjusted R-value is the least, it is evident that the effect of age on the perception of the respondent about the level of data mining within EA is independent. According to the MCA table the respondents with qualification less than a bachelor (mentioned as 'others') with the highest mean score of 4.295 indicating that they are more satisfied with the level of data mining within an environmental accounting in Indian companies. The effect of qualification at the level of data mining within the environmental accounting is found to be independent of the difference between adjusted R-Values 0.294 and unadjusted R- Values 0.128 are comparatively is high. This indicates that the effect of qualification at the level of data mining within the environmental accounting is dependent on other independent variables. The effect of knowledge about Environmental Accounting on the level of data mining within the environmental accounting is found to have an independent effect on the difference between adjusted R-value 0.553, and unadjusted R-Value 0.335 is the highest. Respondents with rich knowledge are more satisfied with the level of data mining within an environmental accounting (4.443). The satisfaction levels of respondents in the other levels of knowledge are founded to be more or less equal. As per the years of experience, it can be seen that the respondent with experience between five to ten years has highest mean opinion score of 3.991 about the level of data mining within Environmental Accounting followed by those with age between eleven to fifteen (3.933). The effect of years of experience on the level of data mining within Environmental Accounting is found to be dependent on the difference between the adjusted R-value (0.120), and unadjusted R-value (0.273) is comparatively higher. The adjusted R-value is lesser than the unadjusted R-value, indicating that the effect of years of experience on the level of data mining was explained away by the other independent variables. It means that the effect of type of years of experience at the level of data mining within the environmental accounting is not independent.

Summary and Conclusion

In our study, these facts confirmed that from the comparison mean scores of by two independent sample t-test in the public and private sector, it could be seen that the market basket analysis and trend analysis are better in private than public sector. The result of two independent sample t-test also shows that accountants in the public and private sector, in comparison of mean scores of objective such fraud detection, market basket analysis, and trend analysis are better in the private sector as the mean scores 4.50, 4.00 and 4.50 respectively which is higher than that of public sector companies. In the terms weight, the most important aspect of determent elements of the level of data mining within public and private sector companies in India respectively belong to; market segmentation, Interactive marketing, trend analysis, market segmentation, customer churn, fraud detection. The two sample t-test display that the accountants have the opinion that the level of data mining is higher within the environmental accounting system as compared to that of the manager. Accountants are more rely on data mining within the environmental accounting system than managers. Also from the comparison of interquartile range (Box-and-Whisker Plot), it can be seen that managers are less consistent than the accountants in relying on data mining within environmental accounting. According to the result of t-test analysis for comparison the opinion of data mining in public and private sector companies; the private sector is more relying on an environmental accounting system for data mining. The results of the multiple regression presented that according to predictor variables like effect of designation, ownership, gender, age, knowledge about environmental accounting, qualification, and experience on the perception of respondents about the level of data mining; the managers' perceptions about the level of data mining within the environmental accounting are lower than that of accountants by an amount of 0.091. Also, all respondents in all age categories have a high perception of the level of data mining within an environmental accounting compared to the reference category (55+years). In terms of knowledge the all respondent categories have a low perception of the level of data mining within an environmental accounting compared to respondents having rich knowledge (reference category). Our study indicated that all qualification categories of respondent are significant in determining the variation in the

perceptions of respondents about the level of data mining within the environmental accounting. The respondents in those categories have a low perception of the level of data mining within an Environmental Accounting compared to less than Bachelor education (reference category). In addition, respondents perception of all experience category is not significant as their significance level is higher than 0.05. Finally; the result indicates that according to the opinion of respondents the level of data mining within environmental accounting is not the same as the mean opinion scores of the respondents of public and private sector companies are different and the effect of the type of company ownership depends upon predictable variables such as designation, age, gender, knowledge about environmental accounting, experience, qualification which varies in public and private sector companies. Hence the result rejects the null hypothesis that 'there is no significant difference existed in the level of data mining between public and private sector industry in India' and this fact confirmed that the level of data mining within environmental accounting is significantly higher in private sector companies in India.

Limitation of Study

The present study was concentrated on public and private sector companies in India, where Environmental Accounting has been in use. The result of the study indicated the need for a control group of companies without using Environmental Accounting to get exhaustive information about the advantages and disadvantages of Environmental Accounting in management function, decision-making and data mining using a comparative study. So by the felt limitation, future researchers are recommended to include a control group of companies without using Environmental Accounting for conducting studies on the similar topic.

References

Yu, T., Lenzen, M., Gallego, B., & Debenham, J. (2006). *A Data Mining System for Estimating a Large size Matrix in Environmental Accounting*. International Conference on Artificial Intelligence Application and Innovation, 2007; Sydney: Australia Faculty of Information Technology, University of Technology, Sydney, NSW.

Alshhadat, A. R., Ghaleb, E. R., & Amer, Q. (2018). The use of data mining techniques in accounting and finance as a corporate strategic tool: An empirical investigation on banks operating in emerging economies. *International Journal of Economics and Business Research*, *15*(4), 442.

Chinchuluun, A., Xanthopoulos, P., Tomaino, V., & Pardalos, P. (2010). Data mining techniques in agricultural and environmental sciences. *International Journal of Agricultural and Environmental Information Systems*, *1*(5), 26–40.

Common, M. (1996). *Environmental and resource economics: An introduction* (2nd ed.). Harlow, Essex: Addison Wesley Longman Ltd Publishers.

Gilberto Carvalho, P., Ricardo, C., & Nelson Francisco Favila, E. (2008). Data mining for environmental analysis and diagnostic: A case study of upwelling ecosystem of Arraial do Cabo. *Azilian Journal of Oceanography*, *56*(3), 1–12.

Peter, R. (2013). Data mining solutions for the business environment. *Database Systems Journal*, *4*(5), 21–29.

Shah, K., Chauhan, P., & Potdar, M. (2014). Data mining techniques in parallel environment - a comprehensive survey. *International Journal of Computer Applications*, *108*(1), 36–41.

Rouse, M. (2017). Data Mining. *TechTarget*. Retrieved August 26, 2018 from https://searchsqlserver.techtarget.com/definition/data-mining

Seifert, J. W. (2004). Data mining and the search for security: Challenges for connecting the dots and databases. *Government Information Quarterly*, *21*(4), 461–480.

Pechenizkiy, M., Puuronen, S., & Tsymbal, A. Why data mining research does not contribute to the business?. In *Proc. of Data Mining for Business Workshop DMBiz (ECML/PKDD'05)*, Porto, Portugal.

Kirkos, E., Spathis, C. H., & Manolopoulos, Y. (2007). Data Mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications*, *32*(4), 995–1003.

Sirikulvadhana, S. (2002). *Data mining as a financial auditing tool*. M.Sc. Thesis in Accounting, Swedish: School of Economics and Business Administration.

Sumathi, S., & Sivanandam, S. N. (2006). *Introduction to data mining and its applications* (2nd ed.). New York, NY: Springer-Verlag Publishers.

Thuraisingham, B. (2000). A primer for understanding and applying data mining. *IEEE Access*, *2*(1), 28–31.

Mendes, R. P., & Vilela J. (2017). Privacy-preserving data mining: methods, metrics, and applications. *IEEE Access*, *5*(9), 10562–10582.

Crespo, F., & Weber, R. A methodology for dynamic data mining based on fuzzy clustering. *Fuzzy Sets and Systems*, *150*(2), 267–284.

Amani, F. M., & Fadlallah, A. (2017). Data mining applications in accounting: A review of the literature and organizing framework. *International Journal of Accounting Information Systems*, *4*(5), 32–58.

Rahman, H. (2008). *Data mining applications for empowering knowledge societies* (2nd ed.). New York, NY: Information Science Reference (an imprint of IGI Global) Publishers.

Data Mining Tutorial: Process, Techniques, Tools & Examples. 2018. Retrieved September 1, 2018 from https://www.guru99.com/data-mining-tutorial.html

Scott, P. D., & Wilkins, E. (1999). Evaluating data mining procedures: Techniques for generating artificial data sets. *Information and Software Technology*, *41*, 579–587.

Ruxandra, P. (2013). Data mining solutions for the business environment. *Database Systems Journal*, *5*(4), 21–29.

Seifert, J. (2004). Data mining and the search for security: Challenges for connecting the dots and databases. *Government Information Quarterly*, *21*, 461–480.