

ethical behavior and CSR activity, without making the connection between the company and its stakeholders. There is no explanation for how the CSR provision fits within the wider ambit of a corporation's role and purpose, the duties expected of its directors, or the information it is expected to disclose.

Until such time that the law is made more precise and backed up by effective enforcement and penalties for non-compliance, it will not promote CSR or make companies engage more with stakeholders. Section 135 is merely a stealth tax and will impose unnecessary compliance burdens.

Conclusion

CSR have no boundaries and are not constrained by race, colour, or religion. Sadly, concern for the community is often mistaken for socialism. On the contrary, every citizen is an asset in economic activity and has opportunities to succeed. CSR is a culture and unwritten contract with the community. This invisible culture can shape brighter future for nations. If employees don't see the point of CSR initiatives, or understand the message, initiatives are unlikely to be effective. Organizations must realize that government alone will not be able to get success in its endeavor to uplift the down trodden of society.

References

- www.csrtimes.com
- <https://csr.gov.in/CSR/index17.php>
- <http://www.mca.gov.in/MinistryV2/csrdatasummary.html>
- https://csr.gov.in/CSR/index_across.php
- <https://www.investopedia.com/terms/c/corp-social-responsibility.asp>
- <https://www.india-briefing.com/news/corporate-social-responsibility-india-5511.html>
- Implementing Corporate Social Responsibility in India: Issues and the Beyond by Ananda Das Gupta
- <http://theconversation.com/in-india-a-legislative-reform-is-needed-to-push-corporate-social-responsibility-80169>
- Business Ethics - Prof. Agalgatti

TECHNIQUES AND APPLICATIONS OF DATA MINING

Prof. Yogesh L. Nangare

Assistant Professor,
Arts, Commerce & Science College Sonai

Abstract

The aim of this paper is to show the process of data mining and how it can help decision makers to make better decisions. It also aims to show the importance of using data mining nowadays. The most important findings are some of the main data mining techniques, algorithms and some of the organizations which have adapted data mining technology to improve their businesses and found excellent results. This review represents the procedure of data mining and how it can be used by any corporate to help the users to get better knowledge from vast amount of data.

1. Introduction

The enlargement of IT sector has produced huge amount of databases and tremendous data in several regions. The study in databases and information technology has set rise to a methodology to save and operate this valuable data for decision making. Data mining is a process of extraction of useful knowledge and patterns from huge data. It is also known as knowledge discovery from data, knowledge mining from data, knowledge extraction or pattern analysis. Data mining is a logical method which is used to examine through huge amount of data so as to find valuable data. The aim of this technique is to find patterns that were formerly unknown. After these patterns are created they can be used to make convinced decisions for development of the businesses.

Three steps involved are:

- Exploration
- Pattern identification
- Deployment

Exploration: In the first step data is cleaned and transformed into another form, and key variables and then nature of data based on the problem are determined.

Pattern Identification: When the data is explored, sophisticated and well-defined for the specific variables the second step is to form pattern identification. Identify and select the patterns which make the finest prediction.

Deployment: Finally the patterns are deployed for desired outcome.

Figure 1 shows the process of Data Mining.

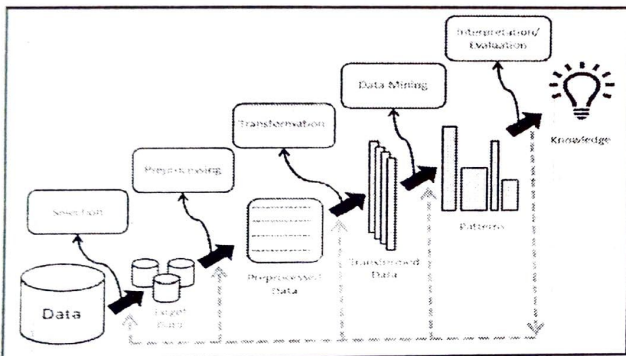


Figure 1. Data Mining Process

2. Data Mining Algorithms and Techniques

Several algorithms and techniques such as Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are used for knowledge extraction from databases.

2.1. Classification

Classification is the most frequently used data mining technique, which provides a set of pre-classified samples to develop a model that can classify the population of records at large. Scam discovery and credit-risk applications are mainly well matched to this type of analysis. This methodology commonly employs decision tree or neural network-based classification algorithms. The data classification technique includes learning and classification. In Learning

the training data are studied by classification algorithm. In classification test data are used to predict the correctness of the classification rules. If the correctness is satisfactory the rules can be applied to the further data tuples. For a scam detection application, this would comprise complete records of both fraudulent and valid activities determined on a record-by-record basis. The classifier-training algorithm uses these pre-classified samples to determine the set of parameters required for appropriate judgment. The algorithm then converts these parameters into a model called a classifier.

Types of classification models:

- Classification by decision tree induction
- Bayesian Classification
- Neural Networks
- Support Vector Machines (SVM)
- Classification Based on Associations

2.2. Clustering

Clustering can be defined as grouping of similar classes of objects. By using clustering techniques we can further recognize dark and light regions in object space and can determine overall distribution pattern and relationships among data attributes. Classification method can also be used for effective means of differentiating groups or classes of object but it becomes expensive so clustering can be used as preprocessing approach for attribute subset selection and classification. For example, to make group of customers based on buying patterns, to categories genes with parallel functionality.

Types of clustering methods

- Partitioning Methods
- Hierarchical Agglomerative (divisive) methods
- Density based methods
- Grid-based methods
- Model-based methods

2.3 Regression

Regression method can be used for prediction. Regression analysis can be applied

to model the relationship among one or more independent variables and dependent variables. In data mining independent variables are elements which are already known and response variables are what we want to predict. Inappropriately, many real-time problems are not just prediction. For example, sales volumes, stock prices, and product failure rates are all very hard to predict as they may depend on complex communications of multiple predictor variables. So, more complex techniques (e.g., logistic regression, decision trees, or neural nets) may be essential to estimate future values. This model types can frequently be used for both regression and classification. For example, the CART (Classification and Regression Trees) decision tree algorithm is used to build both classification trees and regression trees. Neural networks can also create both classification and regression models.

Types of regression methods

- Linear Regression
- Multivariate Linear Regression
- Nonlinear Regression
- Multivariate Nonlinear Regression

2.4. Association rule

Association and connection is frequently find common item set findings between large data sets. This type of discovery supports businesses to make convinced decisions, such as catalogue making, cross advertising and customer shopping activities analysis. Association Rule algorithms need to be capable of generating rules with assurance values less than one. However the a lot of possible Association Rules for a given dataset is usually very huge and a great proportion of the rules are generally of little value.

Types of association rule

- Multi-level association rule
- Multidimensional association rule
- Quantitative association rule

2.5. Neural networks

Neural network is a set of associated

input/output units and each association has a weight present with it. During the training phase, network learns by regulating weights so as to be able to forecast the correct class labels of the input tuples. Neural networks have the outstanding capability to discover meaning from complicated or loose data and can be used to extract patterns and identify trends that are too complex to be recognized by either humans or other computer technology. These techniques are suitable for continuous valued inputs and outputs. For example handwritten character recognition, to train a computer to pronounce English text and many real world business problems and have already been successfully applied in many industries. Neural networks are best for recognizing patterns or trends in data and suitable for prediction or forecasting needs.

Types of neural networks

- Back Propagation
- Convolutional neural network (CNN)
- Multilayer perceptron (MLP)
- Recursive neural network (RNN)
- Recurrent neural network (RNN)

3. Data Mining Applications

Data mining is the process of analyzing a huge amount of data to unlock the hidden information that increases business efficiency. Several businesses have been accepting data mining to their mission-critical business processes to achieve competitive advantages and help business development. This paper explains some data mining applications in sales/marketing, banking/finance, healthcare and insurance, transportation and medicine.

3.1 In Sales/Marketing

Data mining empowers businesses to recognize the hidden patterns inside past purchasing transaction data, thus helping in scheduling and launching new marketing campaigns quickly and profitable way. The following points show several data mining applications in sale and marketing. Data mining is applied for market basket analysis to gain

information on what product combinations were bought together when they were bought and in what sequence. This knowledge assist businesses promote their most moneymaking products and maximize the profit. In addition, it inspires customers to buy related items that they may have been missed or overlooked.

3.2 In Banking / Finance

Various data mining techniques e.g., distributed data mining have been studied, modeled and developed for credit card scam discovery. Data mining is used to recognize customer's reliability by analyzing the data of customer's buying actions such as the data of occurrence of purchase in a period of time, a total economic value of all purchases and when the last purchase was. By analyzing those magnitudes, the relative degree is produced for each customer. The higher is the score, the more relative reliable the customer. To help the bank to maintain credit card customers, data mining is applied. By analyzing the historical data, data mining can help banks guess customers that probable to change their credit card membership so they can plan and launch different special offers to retain those customers. Credit card expenses by customer groups can be recognized by using data mining. The hidden associations between different financial indicators can be revealed by using data mining. From past market data, data mining enables to find stock exchange rules.

3.3 In Health Care and Insurance

The development of the insurance industry totally depends on the capability to transform data into the knowledge, information or intelligence about customers, competitors, and its markets. Data mining is used in insurance industry recently but taken marvelous competitive advantages to the companies who have applied it successfully. Data mining is used in claims analysis for example recognizing which medical procedures are claimed together. Data mining supports to predict which customers will potentially claim new policies. Data mining

enable insurance companies to identify risky customer's activity patterns. Data mining helps to discover deceitful behavior.

3.4 In Transportation

Data mining helps to discover the distribution timetables among warehouses and outlets and analyze loading patterns.

3.5 In Medicine

Data mining empowers to describe patient activities to see incoming office visits. Data mining helps recognize the patterns of fruitful medical therapies for different illnesses.

Conclusions

Data mining has significance regarding finding the patterns, predicting, extraction of knowledge etc., in different business domains. Data mining techniques and algorithms such as classification, clustering etc., helps in recognizing the patterns to make further decisions in businesses to grow. Data mining has wide application area almost in all industries where the data is generated that's why data mining is deliberated as one of the most essential edges in database and information systems and one of the most capable interdisciplinary developments in Information Technology.

References:

1. Jiawei Han and Micheline Kamber (2006), Data Mining Concepts and Techniques, published by Morgan Kaufman, 2nd ed.
2. Dr. Gary Parker, vol 7, 2004, Data Mining: Modules in emerging fields, CD-ROM.
3. Crisp-DM 1.0 Step by step Data Mining guide from <http://www.crisp-dm.org/CRISPWP-0800.pdf>.
4. Customer Successes in your industry from http://www.spss.com/success/?source=homepage&hpzone=nav_bar.
5. <https://www.allbusiness.com/Technology/computer-software-data-management/633425-1.html>, last retrieved on 15th Aug 2010.
6. <http://www.kdnuggets.com/>.
7. <http://www.zentut.com/data-mining/data-mining-applications/>.