Achieving Competitive Advantage through Data Mining

Nadim Rana

Research Scholar (Computer Science), CMJ University, Shillong, Meghalaya

Abstract— Data Mining is the process of identifying novel, valid, useful patterns of data that is non-trivial and easily understandable. Data mining has several advantages to an organization. Helping organizations to gain Strategic advantage is one such application. Strategic advantage also named as competitive advantage is the measure of competence that one organization has over its competitors. Organizations in order to gain strategic advantage have to acquire in depth knowledge about its competitors. In today's business environment data mining is used by several organizations for several business purposes. This research explores the application of data mining in helping them to gain strategic advantage through enabling proper decision making.

Index Terms— Data Mining, Competitive Advantage.

--◆-----------------------------------

1. INTRODUCTION TO DATA MINING

According to Kudyba and Hoptroff (2001, p 36) defines that data mining is the process by which the analysts apply technology to historical data to represent statistically consistent relationships between variables. Analysts use this process to permit the data tell them what is existing in their business rather than testing the rigorous theory validity against the data samples.

Contrary to that Mena (1999, p 42) defined that data mining is the procedure of inventing meaningful and actionable patterns, trends and profiles by transforming through their website data using technologies of pattern recognition such as machine learning, neural networks and genetic algorithms. Data mining is the automatic invention of usable knowledge from the stored server data and leveraging pattern recognition technology.

Data mining is the process of inventing ideas in data whereas other business intelligence methodologies such as ad hoc queries, standard reports and online analytical processing moves in the opposite direction. They initiate with an idea and then collect data to support it defined by Fox (2001, p 174).

Gupta (2006, p 2) defines that data mining is a set of techniques for effective automated invention of previously unknown, novel, valid, understandable and useful patterns in huge databases. The patterns must be actionable so that they may be used in the decision making process of an organization.

Data mining is a collection of techniques and method for analyzing and exploring collection of data in an automatic or semi-automatic way to predict among these data specific hidden or unknown norms, tendencies or associations special systems outcomes the essentials of useful information while reducing the data quantity. Data mining is also the art of acquiring information that is knowledge from data. Therefore data mining is both predictive and descriptive. The predictive techniques are configured to extrapolate new information based on current information while the descriptive techniques are designed to bring out information that is present but buried in a collection of data (Tuffrey, 2011, p 4).

Hand, Mannila and Smyth (2001, p 1-2) defined that data mining is the analysis of observational data sets to predict unsuspected relationships and to summarize the data in novel ways that are both useful and understandable to the data owner.

Data mining is a business process that interacts with other business processes. Specifically a process does not have a beginning and an end and it is ongoing. Data mining initiates with data then through inspires action or analysis informs which in turn creates data that begets more data mining. The practical consequence is that organizations who need to excel at using their data to develop their business do not view data mining as a sideshow. Instead

their business strategy must include gathering data, analyzing data for big term benefit and acting on outcomes. At the same time data mining readily fits with other strategies for understanding customers and markets (Linoff and Berry, 2011, p 1-2).

Finally according to Frawley W J, Shapiro G P and Matheus C J defined that data mining is the non-trivial acquirement of implicit previously unknown and necessarily useful information from data. This enhances several technical approaches such as summarization of data, clustering, finding dependency networks, learning classification rules, predicting anomalies and analyzing alterations. According to Siebes A and Holshemier M (1994) defines that data mining is the search for global patterns and relationships that occurs in huge databases but are hidden among the huge number of data. These relationships denote valuable knowledge about the database and the objects in the database and if the database is a faithful mirror of the real world registered by the database (Frawley W J, Shapiro G P and Matheus C J, Siebes A and Holshemier M (1994) cited in Agarwal and Tayal (2009, p 4).

2. WHAT IS STRATEGIC ADVANTAGE?

The term strategic advantage defines to those marketplace advantages that exert a decisive impact on the likelihood of organization of future success. These advantages frequently are sources of an organization's present and future competitive success relative to other providers of common products. Strategic advantages arouses from both or either of 2 sources such as: 1) strategically necessary external resources which are leveraged and shaped through major external partnerships and relationships; and 2) core competencies which focuses on expanding and building on the internal capabilities of organization (Blazey, 2009, p 62).

Strategic advantage is defined as a superior performance of finance on a specific market which means that organizations has beyond normal returns. An organization has a strategic advantage if it is capable to create several economic values than marginal competitor in its product market (Xu, Tjoa and Chaudhary, 2007, p 1173).

According to Mather (2005, p 30) defined that the strategic advantage is a collection of hard or duplicate competencies, abilities and capacities consisted within an organization that permits it to compete better within the markets that it performs in. These are the desired performance levels that a company is going to require to compete better within its selected markets.

Jones and Tiley (2003, p 16) described that the strategic advantage enhances from several discrete tasks an organization operates in producing, configuring, delivering, marketing and supporting its product. Each of these tasks can contribute to an organization's similar cost position and creates a basis for differentiation.

Similarly Amason (2010, p 9) defined that strategic advantage is defined as the capability of one organization to perform better than its competitors. An organization that performs above the average of their industry has a strategic advantage. Strategic advantage is the major capability or asset that permits one team to best another.

3. DATA MINING TECHNIQUES

The data mining can be gained by classification and prediction, association discovery, predictions, clustering and sequential Patterns.

A. CLASSIFICATION AND PREDICTION:

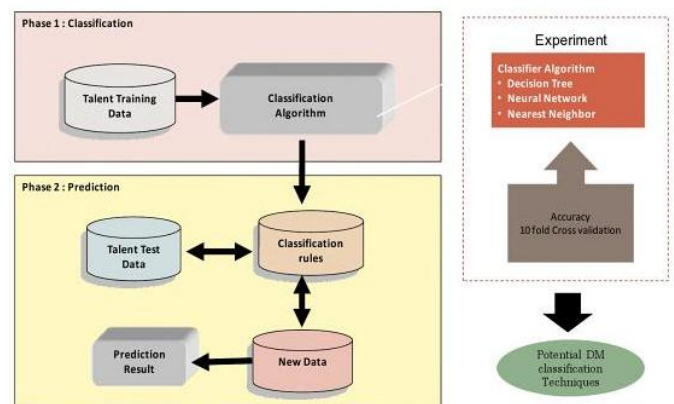The process of classification and prediction in data mining is shown in the below figure:

Figure 1: Classification and Prediction in Data Mining

Source: Setchi R (2010), Knowledge-Based and Intelligent Information and Engineering Systems, Springer, Germany, p 494-495

Classification and prediction capabilities are the methods that can generate intelligent decision. Presently several classification and prediction methods have been proposed by researchers in pattern recognition, statistics and machine learning. Classification and prediction in data mining are 2 forms of data mining techniques that can be used to acquire models to represent necessary data classes or to found future trends of data. In this process the chosen techniques of classification are concerned on similar techniques for prediction and classification

specifically in data mining. The first classification technique selected is neural work which is famous in data mining community and used as the technique of pattern classification. The 2nd technique is decision tree referred to as divide and conquers approach from a collection of independent instances and the 3rd technique is nearest neighbor that is concerned on distance metric.

In the classification process the input variables are academic attributes of talent and the results of this process is academic position. The performance attributes are extracted from the yearly records of performance appraisal, expertise records and previous knowledge. In this process 5 training datasets which consists of seventeen attributes for similar performance components. The classification process has 2 phases such as the 1st phase is learning process whereby training data are analyzed by the algorithm of classification. Classifier or Learned model is represented in the form of classification norms. The 2nd phase is test data and classification are used to represent the classification rules accuracy. If the accuracy is assumed acceptable then the norms can be applied to the new data of classification. The techniques used for classification of data are Bayesian methods, decision tree, neural network, rule based algorithms, association rule mining, support vector machine, case based reasoning, fuzzy logic, k-nearest neighbor, case based reasoning, rough sets and genetic algorithms. The classification techniques consist of neural network, k-nearest neighbor and decision tree. However neural network and decision tree are found useful in developing predictive models in several fields. The decision tree technique benefit is that it does not need any parameter setting or domain knowledge and is proper for exploratory knowledge discovery. The 2nd technique is neural network which has high tolerance of noisy data as well as the capability to categorize pattern on which they have not been trained. Next the k nearest neighbor technique is an instance based learning using distance metric to measure the instances similarity. All these 3 classification techniques have their own benefits and drawbacks (Setchi, 2010, p 494).

B. PREDICTION ANALYSIS:

According to Gopalan and Sivaselan (2009, p 79-80) Prediction Analysis in the data mining technique used to find or appropriately guess the values of data for attributes which are either corrupted or missed. The prediction objective is to make an educated guess on the values of data taking into assumption the values of data possible with other attributes and records in the database. It attempts to find values assuming the format of data distribution in the input database. The data prediction application would be to find the employee's salary in an organization gives the experience details. The whole fact

about prediction is to have predicted value as close as possible to the original or intended value that might have been in place. Prediction is analogous to data classification technique in data mining. The classification of data targets at producing labels of class for records concerned on training data information made possible. The differences lies in the fact that prediction is more oriented towards values of data while classification concentrates on class labels which can be treated as being categorized in nature. A subtle way of differentiating to view the classification as the process of producing overall characteristic of records in terms of prediction and class label information as the process of predicting or producing particular values of data. The prediction of data as a technique incorporates the regression concept from statistics in the predicting values process. Linear regression is used to find values of data for models that are ruled by straight lines while multiple regressions is used to find values for data that do not denote linear properties or are ruled by many variables or predictor. Multiple regressions is also treated as polynomial regressions attempt to transform the non linear data model to a linear one are expansion of their linear counterparts.

C. CLUSTERING:

Pujari (2001, p 54-55) describes that clustering is a method of separating data into various categories so that the data in every group shares common patterns and trends. Clustering forms a main class of DM algorithms. The algorithms attempts to partition the data space automatically into a collection of clusters or regions to which the instances in the table are assigned either probability wise or deterministically. The aim of the process is to recognize all sets of common instances in the data in some optimal fashion. According to similarity clustering is a concept which forms in several disciplines. If the similarity measure is possible then there are several techniques for making clusters. Another approach is to construct a set function that measures some specific group's property. This latter approach gains what is referred as optimal portioning. The clustering objectives are: 1) to initiate data hypothesis; 2) to predict valid and stable organization of the data; and 3) to uncover natural groupings. The clustering methods help to identify various kinds of customers. During the discovery process the differences between data sets can be invented to separate them into various groups and similarity between data sets can be used to group common data together.

D. ASSOCIATION RULE MINING:

According to Nicholson (2009, p 422) described that association rule mining is one of the most famous techniques of data mining used for inventing correlations and associations between data components in a vast

number of applications. The association rule mining first invents all the frequent patterns and then builds the rules from such patterns. Hsu, Lee and Wang (2008, p 116-117) described that the major features of association rule mining is as follows:

On the right hand side no target items: An item can exist on either side of the rule. This is varied from other methods of data mining which typically has some fixed items on the right hand side of the rule as targets.

Completeness: It predicts all possible rules. Other techniques of data mining are only possible to find a subset of rules that exists in data.

Mining with data on hard disk: It does not require loading the entire data into memory. This makes mining a huge dataset possible. Most of the occurring mining techniques need data to be in memory and thus could not manage huge datasets due to major memory size limit.

An association mining algorithm works in 2 steps such as: 1) Produce all frequent item sets that satisfy minsup; and 2) Produce all association rules that satisfy minconf using frequent item sets. An item set is simply a collection of items. Frequent item set that has transaction support above minsup.

E. SEQUENTIAL DATA MINING:

The discovery of sequential pattern endeavors to find ordered patterns such that the set of components presence is followed by another component in a timely common set of episodes or sessions. When browsing a website a user visits various pages in sequence. One may apply sequential mining techniques to invent sequential patterns from web logs. A sequential access pattern indicates the next page most likely to be visited given that the user has visited the previous pages. Mining in the sequential pattern is a necessary technique of data mining with wide applications. Though it is a challenging issue since the mining has to examine or produce a combinatorial wide number of intervening consequences. Present studies have emerged with 2 main sequential pattern mining classes. They are: 1) an applicant test and generation denoted by SPADE which is a format based on vertical mining of sequential pattern mining and GSP which is horizontal format based mining of sequential pattern mining and 2) a sequential method of growth of pattern denoted by CloSpan for closed sequential patterns mining and PrefixSpan for future extensions (Zhong, Liu and Yao (2004, pp 179); Chu and Lin (2005, pp 183).

4. How is data mining helpful in achieving strategic advantage?

Some of the applications of data mining in gaining strategic advantage are: (Laudon & Laudon, 2000, p.53):

I.   Determining which services or products are purchased commonly together such as cigarettes and beer.

II.   Recognizing organization or individuals most likely to answer to direct mailing.

III.   Identifying which transactions are fraudulent.

IV.   Finding which customers switch to rivalries.

V.   Finding what each visitor to a Web site is most interesting in viewing.

VI.   Recognizing similar customer characteristics who buy similar products.

5. ILLUSTRATION

Organizations can identify their strengths and weaknesses by applying data mining techniques using a tool called SAS Enterprise Miner. SAS Enterprise Miner takes as input few attributes that help in determining competitive advantage of organizations and evaluate to what extent organizations are strategically competitive. The tool helps organizations their areas in which they are strong and the areas in which they are weak and guides them the areas that need to be improved in order to gain strategic advantage. The organizations can make use of the results obtained from the tool and apply them in their decision making process and become strategically competitive in the market.

I. SAS ENTERPRISE MINER:

SAS Enterprise Miner is a formidable player in the market of data mining tools. It leverages a necessary influence and power of SAS statistical modules and it enhances that matches by a number of algorithms of data mining. SAS uses its Sample, Explore, Modify, Model and Asses (SEMMA) methodology to provide a tool of data mining that can support a vast number of models including clustering, neural networks, association, statistical regression and decision trees. SAS Enterprise miner is configured to be used both by expert and novice users. Its Graphical User Interface is data flow driven and it is simple to use and understand. It permits an analyst to construct a model by building a visual data flow diagram which links nodes of data with the processing nodes using links. In addition the interface permits for the processing code insertion directly into the flow of data. The Enterprise Miner permits a user to compare various models and chose the best fit by acquiring the assessment node because several nodes are supported. In addition Enterprise Miner offers a scoring

node that generates a scoring model that can be accessed by any SAS application. SAS Enterprise Miner can run in standalone or client/server configurations. Moreover in the client/server mode Enterprise Miner permits the server to be designed as a data server only, compute server or an integration of two. Enterprise Miner is configured to run on all platforms supported by SAS. The architecture supports a client configuration and thin client version (Berson, Smith and Thearling, 2000, p 434).

II. CONCLUSION

It can be understood that the major purpose of data mining is to discovery knowledge that helps in decision making. The tools of data mining find data patterns that they make use of in inferring rules. These norms and patterns can be used to lead the decision-making process of an organization. Data mining helps organizations to speed up the analysis process by paying attention on variables that are most important. The vast reduction in the performance and cost ratio of computer systems has made several organizations to initiate applying composite algorithms used in the techniques of data mining. Today almost all organizations use data mining in their decision making process.

A strategically competitive organization is one that is best capable in anticipating the tactics and strategies of competitors and finds an alternate solution to win. As pointed out by Laudon and Laudon (2000), data mining helps very much in gaining strategic advantage. The analysis using SAS Enterprise miner stands as an evidence for argument by Laudon and Laudon (2003). The data mining process help organizations in identifying their strength and weaknesses and understanding their improvement areas and making better decisions in order to focus towards gaining strategic advantage.

To conclude, several organizations have been implementing data mining projects successfully for a long time. Their success helps continuous evolution and research of data mining. As a result, at present there are several data mining tools in the organization that are critical to classify because of their own benefits and drawbacks. Specifically new tools are being developed consistently to mine huge quantities of data more effectively and offer decision support to organization. It is however, up to the organization to make the best use of them in gaining strategic advantage.

REFERENCES

1. Kudyba and Hoptroff (2001), Data mining and business intelligence: a guide to productivity, Idea Group Publishing, USA, p 36.

2. Mena J (1999), Data mining your website, Digital Press, USA, p 42.

3. Fox J (2001), Building a profitable online accounting practice, John Wiley & Sons, Canada, p 174.

4. Gupta G K (2006), Introduction to Data Mining with Case Studies, PHI Learning Private Limited, New Delhi, p 2.

5. Tuffrey S (2011), Data Mining and Statistics for Decision Making, John Wiley & Sons, UK, p 4.

6. Hand D J, Mannila H and Smyth P (2001), Principles of Data Mining, MIT Press, USA, p 1-2.

7. Linoff G S and Berry M J (2011), Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, John Wiley & Sons, USA, p 1-2.

8. Agarwal B B and Tayal S P (2009), Data Mining and Data Warehousing, Laxmi Publications, New Delhi, p 4.

9. Blazey M L (2009), Insights to Performance Excellence 2009-2010: An Inside Look at the 2009-2010 Baldrige Award Criteria, ASQ Quality Press, USA, p 62.

10. Xu L, Tjoa M and Chaudhary S S (2007), Research and practical issues of enterprise information systems II, Springer, Germany, p 1173.

11. Mather D (2005), The maintenance scorecard: creating strategic advantage, Industrial Press Inc, New York, p 30.

12. Jones O and Tiley F (2003), Competitive advantage in SMEs: organizing for innovation and change, John Wiley & Sons, UK, p 16.

13. Amason A C (2010), Strategic Management: From Theory to Practice, Routledge, New York, p 9.

14. Setchi R (2010), Knowledge-Based and Intelligent Information and Engineering Systems, Springer, Germany, p 494-495.

15. Gopalan N P and Sivaselan B (2009), Data Mining, PHI Learning Private Limited, New Delhi, p 79-80.

16. Pujari A K (2001), Data mining techniques, University Press, Hyderabad, p 54-55.

17.   Nicholson A (2009), AI 2009: Advances in Artificial Intelligence: 22nd Australasian Joint Conference, Melbourne, Australia, December 1-4, 2009, Proceedings, Springer, Germany, p 422.

18.   Hsu W, Lee M L and Wang J (2008), Temporal and spatio-temporal data mining, Idea Group Publishing, New York, p 116-117.

19.   Zhong N, Liu J and Yao Y (2004), Web Intelligence, Springer Verlag, Germany, pp 179.

20.   Chu W W and Lin T Y (2005),"Foundations and advances in data mining", Springer Verlag, New York, pp 183

21.   Kleinbaum D G, Kupper L L and Muller K E (2007), Applied regression analysis and other multivariable methods, Cengage Learning, USA, p 604.

22.   Sullivan L M (2008), Essentials of biostatistics in public health, Jones and Barlett Publishers, Canada, p 63.

23.   Dowdy S, Wearden S and Chilko D (2011), Statistics for Research, John Wiley & Sons, New Jersey, p 495.

24.   Walker G and Shostak J (2010), Common Statistical Methods for Clinical Research with SAS Examples, Third Edition, SAS Publishing, USA.

25.   Berson A, Smith S and Thearling K (2000), Building Data Mining Appl, Tata McGraw Hill, New Delhi, p 434.

26.   Laudon K C and Laudon J P (2000), Management Information Systems: Organization and Technology in the Networked Enterprise, Prentice Hall, New York, p 53.