A Review on Data Mining in Banking Sector

Vikas Jayasree and Rethnamoney Vijayalakshmi Siva Balan

Department of Computer Application,
Noorul Islam University, Kumaracoil, Thuckalay, Kanyakumari (Dt), Tamil Nadu, India

ABSTRACT

The banking industry has undergone various changes in the way they conduct the business and focus on modern technologies to compete the market. The banking industry has started realizing the importance of creating the knowledge base and its utilization for the benefits of the bank in the area of strategic planning to survive in the competitive market. In the modern era, the technologies are advanced and it facilitates to generate, capture and store data are increased enormously. Data is the most valuable asset, especially in financial industries. The value of this asset can be evaluated only if the organization can extract the valuable knowledge hidden in raw data. The increase in the huge volume of data as a part of day to day operations and through other internal and external sources, forces information technology industries to use technologies like data mining to transform knowledge from data. Data mining technology provides the facility to access the right information at the right time from huge volumes of raw data. Banking industries adopt the data mining technologies in various areas especially in customer segmentation and profitability, Predictions on Prices/Values of different investment products, money market business, fraudulent transaction detections, risk predictions, default prediction on pricing. It is a valuable tool which identifies potentially useful information from large amount of data, from which organization can gain a clear advantage over its competitors. This study shows the significance of data mining technologies and its advantages in the banking and financial sectors.

Keywords: Data Mining, Banking Sector, Fraud Detection, Risk Management, Customer Relationship Management

1. INTRODUCTION

The introduction of modern technologies made drastic changes in banking business. The new generation banks with new banking technology and their approaches towards their business, forced other traditional banks to adopt or allocate more focus on new technologies. To improve the financial performance and customer relationship, financial organizations started using web and other electronic channels to process applications for various products, which reduces time and cost. World Wide Web, Electronic and Automated soft-wares have completely changed the basic concepts of banking business and way the business operations are being carried out.

Data Mining is the process of extracting hidden, unknown, valid and actionable information from large databases and then using this information to make crucial business decisions. Previously unknown means quantities that are not hypothesized in advance, valid means if a large collection of data is scrutinized; patterns are not there may be found, Actionable means Action that must be translated into some business advantage (Han *et al.*, 2011). Data mining is the application of statistical and machine-learning techniques for extracting interesting patterns from raw data (Hsu *et al.*, 2012). Data Mining referred as knowledge mining from data or knowledge extraction or data/pattern analysis or data archeology or data dredging. It turns a large collection of data into

Corresponding Author: Vikas Jayasree, Department of Computer Application, Noorul Islam University, Kumaracoil, Thuckalay, Kanyakumari (Dt), Tamil Nadu, India

knowledge (Witten *et al.*, 2011; Liao *et al.*, 2012). With the mounting growth of data in every application, data mining meets the valuable and efficient requirements for effective, scalable and flexible data analysis.

Data Mining is the process of identifying and discovering the interesting patterns from massive amount of data (Mabroukeh and Ezeife, 2010). Data Mining can be conducted on any kind of data as long as the data are meaningful for a target application. Data Mining can be considered as a natural evaluation of information technology and a confluence of several related disciplines and application domains. (Blake and Mangiameli, 2011)

The knowledge Discovery process, depicts as **Fig. 1**, consists of following steps which describes how unprocessed data converts into meaningful information (Bhambri, 2011; Chen *et al.*, 2009):

- Data Selection. In this step, identifies the location of the data and relevance of the data for the business objectives. Because of electronic data are so pervasive, the quality of data plays a critical role in all business and governmental applications (Batini *et al.*, 2009; Yap *et al.*, 2011)
- Data Preparation. Once the data and its location are identified, data cleaning and integration is done in this step. In Data Cleaning, noise data and irrelevant data are removed from the collected data. In Data Integration, different data sources are combined in a common source. Data quality is a major challenge in data mining (Blake and Mangiameli, 2011; Petry and Zhao, 2009) Data analysis underlies many computing applications as a part of their on-line operations or in the design phase. Data analysis procedures can be classified as either exploratory or confirmatory, based on the availability of existing and appropriate models for the data source, but the main point of interest in both types of procedures (whether for hypothesis formation or decision-making) is the grouping, clustering or classification of measurements based on either (i) goodness-of-fit to a postulated model, or (ii) natural groupings (clustering) revealed through analysis (Shinde, 2012)
- Fung *et al.* (2010) Data Transformation. In this step, the selected data is transformed into the form appropriate for the input of data mining process (Han *et al.*, 2011; Moin and Ahmed, 2012; Prakash *et al.*, 2012)
- Data Mining. This is the vital step on which effective algorithms (Bhambri, 2011; Hsu *et al.*, 2012) and techniques applied to process the data into potentially useful patterns to achieve business objectives (Hammawa, 2011; Batini *et al.*, 2009)
- Evaluation. Patterns (Tremblay *et al.*, 2010) representing knowledge are identified based on given

measures (Kontonasios *et al.*, 2012; Wikum *et al.*, 2009)
- Representation. The step on which discovered knowledge visually represents. Visualization techniques (Herawan and Deris, 2011) are more effective in understanding the output for end users

Data Mining encompasses many different techniques and algorithms. These differ in the kinds of data that can be analyzed and the kinds of knowledge representation used to convey the discovered knowledge (Mabroukeh and Ezeife, 2010).

1.1. Application of Data Mining in Banking Sector

There are various areas in which data mining can be used in financial sectors (Ramageri and Desai, 2013; Moradi *et al.*, 2013; Moin and Ahmed, 2012; Hammawa, 2011) like customer segmentation and profitability, credit analysis, predicting payment default, marketing, fraudulent transactions, ranking investments, optimizing stock portfolios, cash management and forecasting operations, high risk loan applicants, most profitable Credit Card Customers and Cross Selling. Certain examples where banking industry has been utilizing the data mining technology effectively as follows.

1.2. Fraud Detection

Fraud detection (Delamaire *et al.*, 2009; Ravisankar *et al.*, 2011; Raj and Portia, 2011; Wang *et al.*, 2009; Petry and Zhao, 2009; Hu and Liao, 2011) is the recognition of symptoms of fraud where no prior suspicion or tendency to fraud exists. According to The American Heritage dictionary, second college edition, fraud is defined as 'a deception deliberately practiced in order to secure unfair of unlawful gain.

Fraud detection refers to detection of criminal activities occurring in commercial organizations such as banks, credit card issuing organizations, insurance agencies, mobile companies, stock market. The malicious users might be the actual customers of the organization or might be posing as a customer (also known as identity theft) (Changdola *et al.*, 2009).

Financial Organizations especially banking sectors follows mainly two approaches (Ramageri and Desai, 2013; Moin and Ahmed, 2012) towards determining the fraud patterns, online transaction check and Offline transaction Check. For this purpose, the institutions purchase and maintain data warehouses of sanctions and Politically Exposed Persons data files from Compliance and Anti Money Laundering solution and data providers like The Office of Foreign Assets Control (OFAC) of the US.
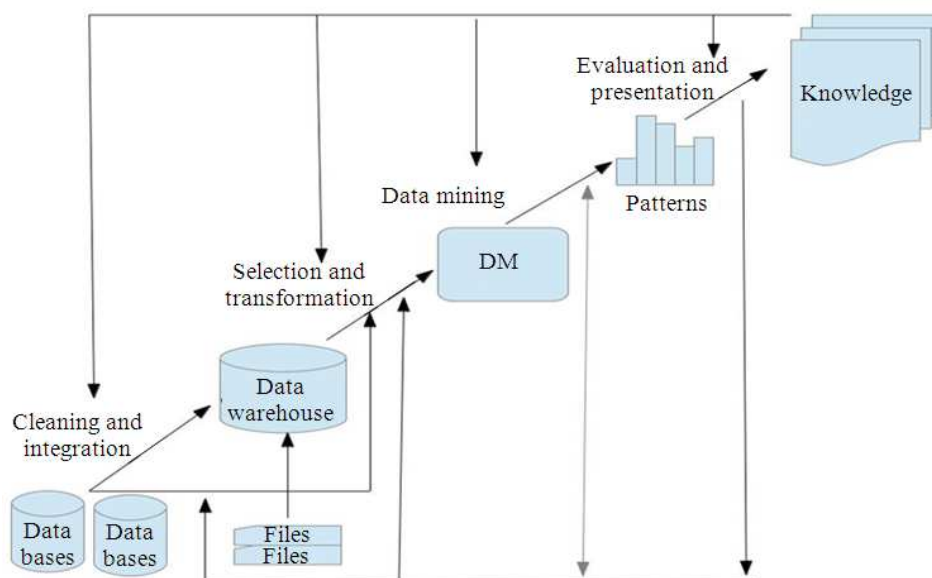
Fig. 1. Data mining as a step in the process of knowledge discovery

Department Treasury, World Check and update upon the changes in external data files. Data mining plays important role in the fraud detection from the transaction data towards external data ware houses. The regulators define certain rules for finding the fraud transaction patterns and data mining process can be applied with different techniques to show the desired fraudulent patterns as output.

Regulatory authorities like Financial Action Task Force (FATAF), Financial Market Supervisory Authority (FINMA), Financial Services Authority (FSA), Hong Kong Monitory Authority (KKMA), Reserve Bank of India (RBI) set standard and requests financial organizations to produce different reports on regular basis. System can be accessed different data sources and preparing reports based on combinations of data mining techniques (Handl and Knowles, 2012; Poovammal and Ponnavaiko, 2009; Wang and Dong, 2009; Issam, 2012; Tremblay *et al.*, 2010; Akbar *et al.*, 2010; Herawan and Deris, 2011; Petry and Zhao, 2009) like classifying, clustering, segmentation, association rules, sequencing, regression, pattern analysis and decision trees.

1.3. Marketing

Most widely used area of data mining in banking technology (Ramageri and Desai, 2013; Wang and Dong, 2009; Hammawa, 2011; Sergio *et al.*, 2011; Bhambri, 2011) is commercial and consumer product marketing. Sales and Marketing department of Financial organizations can use data mining algorithm, to analyze

the existing customers and find the products which they are interested (Petry and Zhao, 2009) and how can they market a another product in association with the existing one. They can use DM techniques to analyze the past trends, find the current demands and predict the customer behavior of various products and services in order to attain more business opportunities, there by establishing or maintaining their position highest in the market (Bhattacharya *et al.*, 2011). Part of maintaining a highest position in the competitive market, financial institution are focusing on promoting unique products with high quality service (Abdullah and Titus, 2010) and its trend analysis can be done by data mining techniques.

Data mining techniques helps strategic planning department to cluster their customers in different buckets like highly potential, good, low and the periodic evaluation on them and thereby providing better service to appropriate clusters (Li *et al.*, 2010). Data mining techniques can be used to identify the customer's reaction on adjustment in interest rates on deposit and borrowing products and its installment changes (Sergio *et al.*, 2011).

Data Mining can improve telemarketing and electronic marketing (Hu and Liao, 2011) by identifying potential customers who are adhere to modern technologies like web, smart phone. In the areas of e-Banking (Aburrous *et al.*, 2009) and other web services used for banking can use another algorithm called sequence pattern-mining effectively. A sequential pattern-mining algorithm mines the sequence database

looking for repeating patterns (Mimno, 2012) (known as frequent sequences) that can be used later by end users or management to find associations between the different items or events in their data (Abdullah and Titus, 2010).

1.4. Risk Management

Data Mining is used to identify the risk factors in each department of banking business (Moradi *et al.*, 2013). Credit Approval authorities in the financial organization used data mining techniques to determine the risk factors in lending decisions (Chen *et al.*, 2009; Chen and Huang, 2011) by analyzing the data based on nationality, repayment capacity and so on. Retail marketing department uses data mining methodologies to find the reliability (Yap *et al.*, 2011) and the behavior of credit card applicant (Delamaire *et al.*, 2009) while selling the credit cards. They uses data mining techniques on existing customers to sell credit cards or increase customers credits or top up on credit card loans (Bhattacharya *et al.*, 2011). In commercial lending, data mining plays a vital role. In commercial lending, risk assessment is usually an attempt to quantify the risk of default or loss to the lender when making a particular lending decision or approving a credit facility (Chen and Huang, 2011).

Here credit risk can be quantified by the changes in the value of a credit products or of a whole credit customer portfolios, which is based on changes in the high risk tendency, default probability, instrument's rating and recovery rate (Yap *et al.*, 2011; Ravisankar *et al.*, 2011; Liu *et al.*, 2012) of the instrument in case of default. The major part of implementation and care of credit risk management system (Raj and Portia, 2011) will be a typical data mining problem: the modeling of the credit instrument's value through the default probabilities, recovery rates and rating migrations (Fung *et al.*, 2010).

Data Mining can be used to derive credit behavior (Delamaire *et al.*, 2009) of individual borrowers with parameters card loans, mortgage value, repayment and using characteristics such as history of credit, employment period and length of residency. A score is thus produced that allow a lender to evaluate the customer and decide whether the person is a good candidate for a loan, or if there is a tendency to become high risk of default (Raj and Portia, 2011). Customers who have been with bank for a longer periods of time, remained better with bank and have good credit history and have higher salaries/wages, are more likely to receive a loan than a new customer who has no credit history with the bank, or who earns low salaries/wages (Ravisankar *et al.*, 2011). Bank can reduce the risk factors to maintain a better position by knowing the chances of a customer to become default (Dorr and Anne, 2009; Tsai, 2012).

1.5. Customer Relationship Management

Data Mining can be useful in all three phases of customer relationship cycle: Customer Acquisition, Increase Value of the customer and Customer Retention (Prakash *et al.*, 2012; Ping and Liang, 2010). Financial organizations especially banking sector recruits Relationship Managers or team of executives to pay proper attention to their customers. Due to the tight competition exists in the market (Sergio *et al.*, 2011; Wang *et al.*, 2009; Chen *et al.*, 2009), customers will always with banks which provide better facility and more secured transaction option. Data Mining techniques (Prakash *et al.*, 2012; Wikum *et al.*, 2009) can be used to determine the list of customers as per the set of definitions (Sergio *et al.*, 2011; Wang *et al.*, 2009; Corne *et al.*, 2012) and interest and the institution can offer better facilities to them (Abdullah and Titus, 2010) customers are varying from their approach in banking, like certain customers interested only electronic banking while others want banking through the counter. Classifying such customers can easily done using data mining techniques and provide better facilities.

Data mining can be used to find out customers holding one product (Wu and Chou, 2011) having interest in similar to other one, there by promoting the product which benefits the organization. Not only can data mining help the banking industry to gain new customers, it can also helps to maintain the existing customers with better service (Tremblay *et al.*, 2010; Kontonasios *et al.*, 2012; Liu *et al.*, 2012).

Within the context of Customer Relationship Management (CRM), data mining can be seen as a business driven process aimed at the discovery and consistent use of profitable knowledge from organizational data (Wu and Chou, 2011). It can be used to fasten the decision making and guide to forecast the effects of decisions (Prakash *et al.*, 2012). Data Mining can be used to increase the response rate of marketing campaign. This can be done by segmenting the customers into groups with their needs and characteristics, it can predict how likely an existing customer is to take the business to a competitor (Mimno, 2012). Each of the CRM elements can be supported by different data mining models (Vaillancourt, 2010; Abdullah and Titus, 2010; Herawan and Deris, 2011; Akbar *et al.*, 2010; Shinde, 2012; Delamaire *et al.*, 2009) which generally include classification, association, sequence discovery, clustering, regression, forecasting and visualization.

2. CONCLUSION

Data Mining is a tool and techniques used to extract meaningful information from the collected data, enables financial institutions to make better decision-making process. Data Collections are in the form of maintaining

proper ware housing based on different databases and other related sources like files into a acceptable data format which becomes the input for data mining process. Based on the standard or rules set by the organization and regulatory authorities, data mining tool extract the knowledge based on the rule set and throws the output in visual tools, thereby making end user life easy to make decisions properly. Banks and Financial organizations started allocating funds and time for implementing data mining tools in the area of decision making by realizing the necessity of data mining in their system.

3. REFERENCES

Abdullah, Z. and H. Titus, 2010. Mining significant least association rules using fast SLP-growth algorithm. Adv. Comput. Sci. Inform. Technol., 6059: 324-336. DOI: 10.1007/978-3-642-13577-4_28

Aburrous, M., M.A. Hussain, K. Dahal and F. Thabatah, 2009. Modelling intelligent phishing detection system for e-banking using fuzzy data mining. Proceedings of the International Conference on Cyber Worlds, Sept. 7-11, IEEE Xplore Press, Bradford, pp: 265-272. DOI: 10.1109/CW.2009.43

Akbar, S., K.N. Rao and J.A. Chandulal, 2010. Intrusion detection system methodologies based on data analysis. Int. J. Comput. Applic., 5: 10-20. DOI: 10.5120/892-1266

Batini, C., C. Cappiello, C. Francalanci and A. Maurino, 2009. Methodologies for data quality assessment and improvement. ACM Comput. Surveys. DOI: 10.1145/1541880.1541883

Bhambri, V., 2011. Application of data mining in banking sector. IJCST, 2: 199-202.

Bhattacharya, S., S. Jha, K. Tharakunnel and J.C. Westland, 2011. Data mining for credit card fraud: A comparative study. Decision Support Syst., 50: 602-613. DOI: 10.1016/j.dss.2010.08.008

Blake, R. and P. Mangiameli, 2011. The effects and interactions of data quality and problem complexity on classification. J. Data Inform. Q. DOI: 10.1145/1891879.1891881

Changdola, V., A. Banerjee and V. Kumar, 2009. Anomaly detection: A survey. ACM Comput. Surveys, 9: 1-72.

Chen, I., L. Chi-Jie, L. Tian-Shyug and L. Chung-Ta, 2009. Behavioral scoring model for bank customers using data envelopment analysis. Stud. Comput. Intell., 214: 99-104. DOI: 10.1007/978-3-540-92814-0_16

Chen, S.C. and M.Y. Huang, 2011. Constructing credit auditing and control and management model with data mining technique. Expert Syst. Applic., 38: 5359-5365. DOI: 10.1016/j.eswa.2010.10.020

Corne, D., D. Clarisse and J. Laetitia, 2012. Synergies between operations research and data mining: The emerging use of multi-objective approaches. Eur. J. Operat. Res., 221: 469-479. DOI: 10.1016/j.ejor.2012.03.039

Delamaire, L., A. Hussein and P. John, 2009. Credit card fraud and detection techniques: A review. Banks Bank Syst., 4: 57-68.

Dorr, D.M. and M.D. Anne, 2009. Establishing relationships among patterns in stock market data. Data Knowl. Eng., 68: 318-337. DOI: 10.1016/j.datak.2008.10.001

Fung, B.C.M., K. Wang, R. Chen and P.S. Yu, 2010. Privacy-preserving data publishing: A survey of recent developments. ACM Comput. Surveys. DOI: 10.1145/1749603.1749605

Hammawa, M.B., 2011. Data mining for banking and finance. Oriental J. Comput. Sci. Technol., 4: 273-280.

Han, J., M. Kamber and J. Pie, 2011. Data Mining Concepts and Techniques. 3rd Edn., Elsevier, Burlington, ISBN-10: 9780123814807, pp: 744.

Handl, J. and J. Knowles, 2012. Clustering criteria in multiobjective data clustering. Parallel Problem Solving Nat., 7492: 32-41. DOI: 10.1007/978-3-642-32964-7_4

Herawan, T. and M.M. Deris, 2011. A soft set approach for association rules mining. Knowl.-Based Syst., 24: 186-195. DOI: 10.1016/j.knosys.2010.08.005

Hsu, F.M., L.P. Lu and C.M. Lin, 2012. Segmenting customers by transaction data with concept hierarchy. Expert Syst. Applic., 39: 6221-6228. DOI: 10.1016/j.eswa.2011.12.005

Hu, Y.C. and P.C. Liao, 2011. Finding critical criteria of evaluating electronic service quality of internet banking using fuzzy multiple-criteria decision making. Applied Soft Comput., 11: 3764-3770. DOI: 10.1016/j.asoc.2011.02.008

Issam, A.R.M., 2012. Incremental dependency aided metapattern generator. Malaysian J. Comput. Sci., 25: 149-163.

Kontonasios, K.N., E. Spyropoulou and T.D. Bie, 2012. Knowledge discovery interestingness measures based on unexpectedness. Wiley Interdisciplinary Rev.: Data Min. Knowl. Discovery, 2: 386-399. DOI: 10.1002/widm.1063

Li, W., X. Wu, Y. Sun and Q. Zhang, 2010. Credit card customer segmentation and target marketing based on data mining. Proceedings of the International Conference on IEEE Computational Intelligence and Security (CIS), Dec. 11-14, IEEE Xplore Press, Nanning, pp: 73-76. DOI: 10.1109/CIS.2010.23

Liao, S.H., P.H. Chu and P.Y. Hsiao, 2012. Data mining techniques and applications-A decade review from 2000 to 2011. Expert Syst. Applic., 39: 11303-11311. DOI: 10.1016/j.eswa.2012.02.063

Liu, B., Y. Li and Y.C. Tian, 2012. Discovering novel knowledge using granule mining. Rough Sets Curr. Trends Comput., 7413: 380-387. DOI: 10.1007/978-3-642-32115-3_45

Mabroukeh, N.R. and C.I. Ezeife, 2010. A taxonomy of sequential pattern mining algorithms. ACM Comput. Surveys. DOI: 10.1145/1824795.1824798

Mimno, D., 2012. Computational historiography: Data mining in a century of classics journals. J. Comput. Cultural Heritage.

Moin, K.I. and Q.B. Ahmed, 2012. Use of data mining in banking. Int. J. Eng. Res. Applic., 2: 738-742.

Moradi, M., M. Salehi, M.E. Ghorgani and H.S. Yazdi, 2013. Financial distress prediction of Iranian companies by using data mining techniques. Organizacija, 46: 20-27.

Petry, F.E. and L. Zhao, 2009. Data mining by attribute generalization with fuzzy hierarchies in fuzzy databases. Fuzzy Sets Syst., 160: 2206-2223. DOI: 10.1016/j.fss.2009.02.014

Ping, Z.L. and S.Q. Liang, 2010. Data mining application in banking-customer relationship management. Proceedings of the IEEE International Conference on Computer Application and System Modelling, Oct. 22-24, IEEE Xplore Press, Taiyuan, pp: 124-126. DOI: 10.1109/ICCASM.2010.5619002

Poovammal, E. and M. Ponnavaiko, 2009. An improved method for privacy preserving data mining. Proceedings of the IEEE International Advance Computing Conference, Mar. 6-7, IEEE Xplore Press, Patiala, pp: 1453-1458. DOI: 10.1109/IADCC.2009.4809231

Prakash, B.V.A., D.V. Ashoka and V.N.M. Aradhya, 2012. Application of data mining techniques for software reuse process. Proc. Technol., 4: 384-389. DOI: 10.1016/j.protcy.2012.05.059

Raj, S.B.E. and A.A. Portia, 2011. Analysis on credit card fraud detection methods. Proceedings of the International Conference on Computer Communication and Electrical Technology, Mar. 18-19, IEEE Xplore Press, Tamilnadu, pp: 152-156. DOI: 10.1109/ICCCET.2011.5762457

Ramageri, B.M. and B.L. Desai, 2013 Role of data mining in retail sector. Int. J. Comput. Sci. Eng., 5.

Ravisankar, P., V. Ravi, G.R. Rao and I. Bose, 2011. Detection of financial statement fraud and feature selection using data mining techniques. Decision Support Syst., 50: 491-500. DOI: 10.1016/j.dss.2010.11.006

Sergio, M., R. Laureano and C. Paulo, 2011 Using data mining for bank direct marketing: An application of the CRISP-DM methodology. Proceedings of the European Simulation and Modelling Conference, (ESMC' 11), Guimarães, Portugal, pp: 117-121.

Shinde, P., 2012. Data mining using artificial neural network tree. IOSR J. Eng.

Tremblay, M.C., K. Dutta and D. Vandermeer, 2010. Using data mining techniques to discover bias patterns in missing data. J. Data Inform. Q., 2: 1-19. DOI: 10.1145/1805286.1805288

Tsai, H.H., 2012. Global data mining: An empirical study of current trends, future forecasts and technology diffusions. Expert Syst. Applic., 39: 8172-8181. DOI: 10.1016/j.eswa.2012.01.150

Vaillancourt, J., 2010. Statistical methods for data mining and knowledge discovery. Proceedings of the 8th International Conference on Formal Concept Analysis, Mar. 15-18, Springer Berlin Heidelberg, Agadir, Morocco, pp: 51-60. DOI: 10.1007/978-3-642-11928-6_4

Wang, G., G. Nie, P. Zhang and Y. Shi, 2009. Personal financial market segmentation based on clustering ensembles. Proceedings of the WRI World Congress on Computer Science and Information Engineering, Mar. 31-Apr. 2, IEEE Xplore Press, Los Angeles, CA., pp: 694-698. DOI: 10.1109/CSIE.2009.741

Wang, X. and G. Dong, 2009. Research on money laundering detection based on improved minimum spanning tree clustering and its application. Knowl. Acquisit. Model., 2: 62-64. DOI: 10.1109/KAM.2009.221

Wikum, G., G. Kasneci, M. Ramanath and F. Suchanek, 2009. Database and information-retrieval methods for knowledge discovery. Mag. Commun. ACM, 52: 56-64. DOI: 10.1145/1498765.1498784

Witten, I.H., E. Frank and M.A. Hall, 2011. Data Mining: Practical Machine Learning Tools and Techniques. 3rd Edn., Morgan Kaufmann, ISBN-10: 0123748569, pp: 664.

Wu, R.S. and P.H. Chou, 2011. Customer segmentation of multiple category data in e-commerce using a soft-clustering approach. Elect. Commerce Res. Applic., 10: 331-341. DOI: 10.1016/j.elerap.2010.11.002

Yap, B.W., S.H. Ong and N.H.M. Hussain, 2011. Using data mining to improve assessment of credit worthiness via credit scoring models. Expert Syst. Appli., 38: 13274-13283. DOI: 10.1016/j.eswa.2011.04.147