



Literature Review for Data Mining and Industrial Applications

Tuğba Efendigil¹, İnci Elif Sağlam^{1*}

Abstract

Along with rapid improvement in data collection and storage technology, organizations have faced with vast amount of data. However, extracting useful and significant information has presented extremely challenging. For this reason, data mining techniques including generalization, characterization, classification, clustering, association, evolution, pattern matching, and data visualization increase in importance for solving a variety of industry problems. Data mining facilitates to search a huge database without any specific pre-determined hypothesis to obtain implicit, previously unknown, and potentially useful information. In literature, data mining techniques are applied a variety of area such as finance to detect fraud or risk to early warning system, customer relationship management (CRM) to analyze customer profile and behavior, manufacture to predict defections and quality improvement, banking to forecast bankruptcy, marketing and direct marketing, healthcare, information systems and so on. This study aims to reveal an overview about application of data mining techniques occurred in literature. Thus, a wide perspective related to various data mining techniques and its applications is gained for the researchers who are willing to study of future works.

Keywords: Big Data, Data Mining, Data Mining Applications, Data Mining Techniques

Introduction

In last decades, data gathering and storing technology has progressed rapidly. Although the collected huge amount of data includes considerable information, interpretation of the data is getting hard day by day. Data mining has emerged to benefit from increasing information to better understand the markets, customers, suppliers, operations, internal business processes, patients, students/teachers and so forth. It provides valid, comprehensible, and potentially useful knowledge from large data sets and thus facilitates decision making process. Data mining integrates concepts from modern statistics, intelligent information systems, machine learning, pattern recognition, decision theory, data engineering, and database management[1]. It provides beneficial tools to find out complex and hidden relationships in large amounts of data.

Data mining contains several algorithms that serve a variety of purposes such as prediction, description, and association etc. Classification techniques, especially decision tree, neural network (NN), support vector machine (SVM), naïve bayes, clustering, regression models, association rule algorithms are widely preferred to fulfill the goals.

In literature, there are several data mining applications which are preferred for different areas such as CRM, medicine, finance and banking, manufacturing, education, marketing, information technology etc. This study focuses on four main areas as; finance/banking, CRM, manufacturing and marketing. Within this context, 260 published papers in English were reviewed which have been presented in several journal data bases like Science Direct, Scopus, IEEE Xplore, Wiley, Springer Link, and Emerald. The review, which includes conference proceedings and journal publications, states a general outlook as follows.

¹ Tuğba EFENDİGİL, Yıldız Technical University, Mechanical Faculty, Department of Industrial Engineering, Istanbul, Turkey, tugbae@yildiz.edu.tr

^{1*} İnci Elif SAĞLAM, Yıldız Technical University, Mechanical Faculty, Department of Industrial Engineering, Istanbul, Turkey, iesaglam@yildiz.edu.tr



Literature Review

Although a few data mining reviews have been provided in different times, all they referred only one aspect of data mining. Ngai et al. (2011) [2] made a review related to financial fraud detection, Rygielski et al. (2002) [3] and Ngai et al. (2009) [4] touched on data mining techniques used to CRM, Choudhary et al. (2009) [5] and Köksal et al. (2011) [6] tackled manufacturing applications and quality improvement in manufacturing. There is not any literature survey related to industrial applications of data mining. To fill the gap, this study gives an insight about data mining implementations in industries.

Even though first application was made in 1993, most of data mining studies belong to 2000s. Number of these studies have increased in progress of time and reached maximum number in between 2004-2006 years. CRM is becoming a widespread application area while it is followed by manufacturing and finance. In recent years, with the development in service industry, marketing applications have accelerated along with CRM.

This section presents a detailed literature review related to data mining and its applications. The research is arranged based on practices of data mining in various industries such as finance and banking, CRM, manufacture and marketing respectively. Moreover different data mining techniques like that Artificial Neural Network (ANN), Decision Tree (DT), CHAID Decision Tree and Logistic Regression are discussed in detail in terms of their usage rate.

Finance and Banking Sector

Data mining has already a major impact on business and finance. Financial markets generate large volumes of data. Analyzing these data to reveal valuable information and making use of the information in decision making present great opportunities and grand challenges for data mining[1]. A great variety of data mining techniques has been widely applied to solve problems in finance fields, including financial early warning system[7, 8], fraud detection, financial distress and bankruptcy prediction[9], financial performance estimation[10, 11], credit risk management[12, 13], bank lending, and so on.

Data mining techniques have been widely used to accomplish task of management fraud detection in credit card, insurance fraud and general fraud problem. In literature there are some researches that giving an insight about the use of data mining techniques[2, 14] and investigate usefulness of data mining techniques in fraud detection area [15]. Kotsiantis et al. [16] used a variety of techniques including Artificial Neural Network (ANN), Decision Tree (DT), Logistic Regression, Bayesian Network, Support Vector Machine (SVM) to forecast fraudulent financial system and compare performance of these techniques with together. Similarly, Ravisankar et al.[17]utilize Multilayer Feed Forward Neural Network (MLFF), Support Vector Machines (SVM), Genetic Programming (GP), Group Method of Data Handling (GMDH), Logistic Regression (LR), and Probabilistic Neural Network (PNN) techniques to identify companies that resort to financial statement fraud. Neural Network [18-20], Classification and Regression Tree (CART) [21], and text mining techniques [22] are preferred to detect fraud. Especially credit card fraud is serious and growing problem so detection of this fraud has gained importance. Decision Tree (DT) [23, 24], Neural Network (NN) [25, 26], Self-Organization Map (SOM) [27, 28], and Support Vector Machine (SVM) and Logistic Regression [29] techniques are frequently preferred to serve this purpose. Also, commentators within the insurance sector have also pointed to the difficulties in estimating the prevalence and costs of insurance fraud[30]. Thus, disclosure of this fraud has become important for academic researches and several data mining techniques are applied to a variety of insurance sectors such as automobile[31-36], healthcare[37], medical insurance fraud [38]etc.

Prediction of financial distress has been a topic of interest over the decades because of its great importance to listed companies, interested stakeholders and even the economy of a country[39]. If financial distress is predicted effectively, managers of companies can initiate proactive measures to avoid deterioration before the crisis. Thus investors can grasp the profitability situation of the listed companies and adjust their investment strategies to reduce anticipated investment related losses. Support Vector Machine (SVM)[40-42] and Radial Basis Function Support Vector Machine (RSVM)[43], Neural Network (NN) [44, 45] and Probabilistic Neural Network [46], Decision Tree (DT) [47, 48] and Ada Boosted Decision Tree [49] and k-Nearest Neighbors Techniques [50] are widely used in literature to predict and analyze financial distress. In addition to estimating of financial distress, predicting bankruptcy [51-59] and financial performance [60-62] has been applied as well.



Customer Relationship Management (CRM)

Many organizations have collected and stored numerous data about their current customers, potential customers, suppliers and business partners. However, the inability to discover valuable information which is hidden in data prevents the organizations from transforming these data into valuable and useful knowledge [63]. Data mining tools could help the organizations discover the hidden knowledge in this enormous data.

Analyzing and understanding of customer behaviors and profiles is foundation of development for a competitive CRM strategy, so as to acquire and maintain potential customers and maximize customer value. Hence, application of data mining in CRM is an emerging trend in global market and academic researches. Particularly data mining enables the extraction of hidden predictive information from large databases, thus organizations can identify valuable customers, predict future behaviors, and make companies proactive with knowledge-driven decisions.

According to Kracklauer et al. [64] CRM consists of four dimensions such as customer identification, customer attraction, customer retention and customer development.

- *Customer identification:* CRM begins with customer identification that involves targeting the population who are most likely to become customers or most profitable to the company. Elements for customer identification include customer segmentation, target customer analysis and customer profile/behavior analysis. Data mining is a widely used technique in customer identification under different focus points with different techniques. In customer segmentation area there are a large number of publications including decision tree [65, 66], SOM and Markov Chain [67-70], SOM and NN [71, 72], regression [73], k-means [74], k-means and SOM [75-77], k-means and rough set theory [78], k-means and association rule [79], SOM and fuzzy clustering and NN [80] and other clustering techniques [81-84]. Decision tree [85-87], SOM [88] and k-means [89] are preferred for target customer analysis tasks. Finally, in customer profile/behavior analysis decision tree [90, 91], decision tree and sequential pattern mining [92, 93], decision tree, naïve bayes and SVM [94, 95], backpropagation NN [96] and fuzzy clustering [97] are widely utilized techniques.
- *Customer attraction:* Customer attraction is the phase following customer identification [4]. After identifying the segments of potential customers, organizations can give the effort and resources into attracting the target customer segment. A component of customer attraction is direct marketing which is a promotion process motivating customers to place orders through various channels [98-101]. A vast number of researches that applied data mining techniques such as logistic regression [101], NN [102-104], decision tree [105] and genetic algorithm [106, 107] are presented in literature.
- *Customer retention:* Customer retention is the central concern for CRM. Customer satisfaction, which refers to the comparison of customers' expectations with the perception of being satisfied, is the essential condition for retaining customers [64]. Customer retention includes loyalty programs, one-to-one marketing, and complaints management [4]. Loyalty programs involve campaigns or supporting activities which aim at maintaining a long term relationship with customers. Data mining techniques, especially decision tree [108-111], SOM [112, 113], SOM and association rule mining [114], logistic regression [115, 116], NN, decision tree and k-nearest neighbor [117], genetic algorithm [118], k-means [119] are preferred to analyze and improve customer loyalty. Specifically, churn analysis, credit scoring, service quality or satisfaction constitutes part of loyalty programs. Under those subtitles, logistic regression [120-124], logistic regression and some combined techniques such as SVM [120, 125], NN and decision tree [126, 127] are widely used. In addition to these techniques, decision tree [128], decision tree and NN [129], NN and SOM [130] are preferred. One-to-one marketing refers to personalized marketing campaigns which are supported by analyzing, detecting and predicting changes in customer behaviors [131]. Association rule mining (ARM) [131-140], ARM and SOM [67, 141], ARM and NN [142] are frequently applied to one-to-one marketing implementation. SOM [143-145], SVM [98], decision tree [146], NN [147, 148], nearest neighbor



[149] are also used to fulfill the purposes. A variety of data mining techniques as SOM [150], ARM [151, 152], SVM [153] are referred to complaints management which is another inseparable part of customer retention.

- *Customer development:* Customer development involves consistent expansion of transaction intensity, transaction value and individual customer profitability [4]. Customer lifetime value analysis, value analysis, up/cross selling and market basket analysis are components of customer development. For life time value analysis tasks, classification techniques, particularly NN [154], NN and decision tree [155], SVM [156], nearest neighbor and ARM [157], linear regression [158], bayesian network classifier [159] are applied. As distinct from life time value analysis, customer value refers to the potential customer contributions to an enterprise during specific periods. There are a few publications [160-162] related to value analysis in comparison with the life time value analysis. ARM [163] and hybrid model including ARM [164, 165] most popular techniques to up/cross selling implementation. Similar to up/cross selling phase, ARM [166-170] is overly applied to market basket analysis.

Manufacture

In modern manufacturing systems, huge amounts of data are stored in database management systems and data warehouses from all involved areas such as product and process design, assembly, materials planning, quality control, scheduling, maintenance, fault detection etc. [5]. Data mining techniques have been utilized in manufacture environment to analyze the enormous data and getting useful information.

This part follows on literature research based on six categories that are widely mentioned in literature.

- *Product design:* Kusiak and Smith [171] focused on product design and its data mining applications providing a variety of industrial cases and scenarios related to clustering techniques. K-means [172] and c-means clustering [173] are extensively preferred as clustering techniques. Also, some classification techniques such as neural network [174-176], decision tree [177, 178] and association rule analyze techniques [179] and sequential pattern [180] are used to product design process. In addition to common data mining techniques text mining [181] and semantic data mining [182] are applied to component design.
- *Manufacturing process design, monitoring and control:* Relationship among machines may be necessary to design of production process. Chen [183] found certain sets of machines by using association rule and developed an effective configuration of cellular manufacturing. Also Cunha et al. [184] applied association rule to make assembly schedule. Monitoring and control and defect detection have vital importance to manufacturing process. Monitoring and control help to analyze abnormal condition and avoid loss. Along with advance in monitoring technology, automated monitoring and control system is adapted, thus analyzing of this process has become so complicate and difficult in time. Artificial neural network, genetic algorithm [185], support vector machine [186], self-organizing map [187], regression analysis [188], rough set theory [189] c-means clustering [190] are applied manufacture data to aim of monitoring and controlling. In addition to these, association rule mining [191] and clustering techniques like k-means [192] and k-nearest neighbor [193] are utilized to detect certain product defect.
- *Job-shop scheduling:* Job shop scheduling is an important and complex activity in manufacturing. A job shop model, typical in manufacturing, can be described as a set of jobs composed of sequences of operations that are processed on a set of machines. It is a decision making process which allocates limited resources over time to perform a set of jobs to meet objectives [194]. Decision tree [195], regression tree [196], neural network [197], rough set theory [198] and genetic algorithm [194, 198] are used to make effective decision.
- *Fault diagnosis:* As data sets increase in size, exploration, manipulation and analysis become more complicated and resource consuming to evaluate probability of failure data relating to environment and operation condition. To deal with these situations identifying classes of process faults decision tree



[199-201] and support vector machine [202, 203], rough set theory and neural network [204] and k-medoids and Gustafson-Kessel algorithm [205] are widely used.

- *Quality monitoring, analysis and improvement:* Quality improvement of industrial products and processes requires collection and analyses of data to cope with quality related manufacturing problems. [6]. Owing to advances in data collection systems and analysis tools, data mining (DM) has widely been applied the quality problem in manufacturing. Neural network [206], association rule mining [207], support vector machine and k-means [208] for quality improvement, fuzzy clustering techniques [209] for quality improvement, back propagation neural network [210], self-organizing map, k-means, decision tree [211] for quality prediction, decision tree, k-nearest neighbor and naïve bayes[212] for quality control, neural network [213] for quality performance assessing are used extensively.
- *Yield improvement:* Data mining has proved to be important for yield improvement which is a key element to maintain efficiency of semiconductor manufacturing[5]. A variety of classification techniques such as decision tree [214],[215, 216], hybrid form decision tree with neural network [217, 218], neural network [219-221] and hybrid form neural network with spatial statistics [222], regression analysis [223], genetic programming [224], a combination of neural network, self-organizing map and pattern recognition [225], decision tree and sequential pattern [226] are preferred in semiconductor manufacture to improve yield. In addition to semiconductor sector, decision tree technique [227],[228] is used to enhance productivity in other electronic manufacture sector.

Marketing

Marketing is one of leading areas where data mining techniques have been applied. Data mining enables organizations to sort through vast amounts of customer data by discovering purchase behavior of customer [67], [229]pattern of price information [230], interest group [231] and forecast sales [232], [233], [234], demand [235] and product availability [236] and so on. This is a vital importance for marketing departments of any organization in any sector. If an organization knows what their customers require and are able to predict what their spending patterns will be, remarkable amounts of time and money can be saved[237].

Understanding customers' buying habits and identifying their profiles are essential in marketing. Data mining can provide that information[237]. K-means and apriori algorithm are extensively used together to discover customer profile [238], to mine customer knowledge [239], [240], [241], to discover purchase pattern [242], to analyze return pattern [243]and store segmentation [244]. Also, other clustering techniques such as hierarchical clustering with association rule [245], CAVE clustering algorithm to identify similarity between categorical values [246], fuzzy c-means with neural network to explore customer profile [247] are preferred. Market basket analysis [248] is utilized on the purpose of segmentation. In addition to these, as we mentioned in CRM part, the general incidence of data mining use for direct marketing[249],[239], [250], cross-selling[251], [252]and one-to-one marketing[253]has grown day by day.

According to Chen et al. [254]recent marketing research has suggested that in-store environmental stimuli, such as shelf-space allocation and product display, has a great effect on consumer buying behavior and may induce substantial demand. And also retail industry is also realizing that it is possible to gain a competitive advantage utilizing data mining. Association rule mining technique [255], [254], [256], [167, 257] and sequential pattern [258]are frequently preferred to self-space allocation in retail. Market basket analysis[259], [167], [260] is a marketing method used by many retailers to determine the optimal locations to promote products and develop efficient store layout.

Conclusion

Aim of this review is to give an insight about data mining applications in different industries. Popular areas and applications are presented and gaps are discovered. In our research, 260 studies are viewed by scanning several journal databases. Although first application made in 1993, most of data mining studies belong to 2000s. Number of these studies has increase in progress of time and reached maximum number in between 2004-2006 years. According to the literature research, CRM is the most popular area over others in every year by being involved in 108 studies after 2000. CRM is followed by manufacturing which many of studies were



published in the middle of 2000s. Earliest manufacturing application was presented in 1993 and from that time 60 studies related to manufacturing have been appeared in literature. Contrary to CRM and manufacturing, finance and banking applications have flourished during the recent years. Totally, 57 publications are found and the first study in this area was written in 1997. Lastly, marketing applications have increasingly developed in time and 40 papers were submitted in between 1997-2014. When examining the data mining techniques; classification techniques, especially, decision tree, neural network, support vector machine, naïve bayes are more preferred in comparison with other techniques. Also, hybrid models that are used for classification, clustering and rule based techniques together, have come in to prominence.

As a conclusion, mining of data is a young field to be improved by both industries and universities. Especially dealing with big data analytics will help organizations learn the expectations and satisfaction of their customers via their feedbacks and behaviors. After investigating many publications, marketing is determined to have a potential as an emerging trend for data mining while considering the profitability of its implementations and the virginity in academic approaches.

References

- [1] H. Wang and A. S. Weigend, "Data Mining for Financial Decision Making," *Decision Support Systems*, vol. 37, pp. 457-460, 2004.
- [2] E. W. T. Ngai, Y. Hu, Y. H. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decision Support Systems*, vol. 50, pp. 559-569, 2011.
- [3] C. Rygielski, J.-C. Wang, and D. C. Yen, "Data mining techniques for customer relationship management," *Technology in Society*, vol. 24, pp. 483-502, 2002.
- [4] E. W. T. Ngai, L. Xiu, and D. C. K. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification," *Expert Systems with Applications*, vol. 36, pp. 2592-2602, 2009.
- [5] A. K. Choudhary, J. A. Harding, and M. K. Tiwari, "Data mining in manufacturing: a review based on the kind of knowledge," *Journal of Intelligent Manufacturing*, vol. 20, pp. 501-521, 2009.
- [6] G. Köksal, İ. Batmaz, and M. C. Testik, "A review of data mining applications for quality improvement in manufacturing industry," *Expert Systems with Applications*, vol. 38, pp. 13448-13467, 2011.
- [7] L. Zhang, L. Zhang, W. Teng, and Y. Chen, "Based on InformationFusion Technique with Data Mining in the Application of Finance Early-Warning," *Procedia Computer Science*, vol. 17, pp. 695-703, 2013.
- [8] W. L. Tung, C. Quek, and P. Cheng, "GenSo-EWS: a novel neural-fuzzy based early warning system for predicting bank failures," *Neural Networks*, vol. 17, pp. 567-587, 2004.
- [9] C. Sevim, A. Oztekin, O. Bali, S. Gumus, and E. Guresen, "Developing an early warning system to predict currency crises," *European Journal of Operation Research*, vol. 237, pp. 1095-1104, 2014.
- [10] T. Y. Kim, K. J. Oh, I. Sohn, and C. Hwang, "Usefulness of artificial neural networks for early warning system of economic crisis," *Expert System with Applications*, vol. 26, pp. 583-590, 2004.
- [11] S.-H. Liao and S.-Y. Chou, "Data mining investigation of co-movements on the Taiwan and China stock markets for future investment portfolio," *Expert Systems with Applications*, vol. 40, pp. 1542-1554, 2013.
- [12] A. S. Koyuncugil and N. Ozgulbas, "Financial early warning system model and data mining application for risk detection," *Expert System with Applications*, vol. 6, pp. 6238-6253, 2012.
- [13] A. Khashman, "Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes," *Expert Systems with Applications*, vol. 37, pp. 6233-6239, 2010.
- [14] G. L. Gray and R. S. Debreceeny, "A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits," *International Journal of Accounting Information Systems*, vol. 15, pp. 357-380, 2014.
- [15] E. Kirkos, C. Spathis, and Y. Manolopoulos, "Data Mining techniques for the detection of fraudulent financial statements," *Expert Systems with Applications*, vol. 32, pp. 995-1003, 2007.
- [16] S. Kotsiantis, E. Koumanakos, D. Tzelepis, and V. Tampakas, "Forecasting fraudulent financial statements using data mining," *International Journal of Computational Intelligence*, vol. 3, pp. 104-110, 2006.
- [17] P. Ravisankar, V. Ravi, G. Raghava Rao, and I. Bose, "Detection of financial statement fraud and feature selection using data mining techniques," *Decision Support Systems*, vol. 50, pp. 491-500, 2011.
- [18] K. M. Fanning and K. O. Cogger, "Neural network detection of management fraud using published financial data."
- [19] M. J. Cerullo and V. Cerullo, "Using neural networks to predict financial reporting fraud: Part 2," *Computer Fraud&Security*, vol. 1999, pp. 14-17, 1999.
- [20] J. W. Lin, M. I. Hwang, and J. D. Becker, "A fuzzy neural network for assessing the risk of fraudulent financial reporting," *Managerial Auditing Journal*, vol. 18, pp. 657-665, 2003.



- [21] B. Bai, J. Yen, and X. Yang, "False Financial Statements: Characteristics of China's Listed Companies and Cart Detecting Approach," *International Journal of Information Technology & Decision Making*, vol. 07, pp. 339-359, 2008.
- [22] C. Holton, "Identifying Disgruntled Employee Systems Fraud Risk Through Text Mining: A Simple Solution for a Multi-billion Dollar Problem," *Decision Support Systems*, vol. 46, pp. 853-864, 2009.
- [23] P. K. Chan, F. Wei, A. L. Prodromidis, and S. J. Stolfo, "Distributed data mining in credit card fraud detection," *IEEE Intelligent Systems*, pp. 67-74, 1999.
- [24] I. G. N. N. Mandala, C. B. Nawangpalupi, and F. R. Praktikto, "Assessing Credit Risk: An Application of Data Mining in a Rural Bank," *Procedia Economics and Finance*, vol. 4, pp. 406-412, 2012.
- [25] J. R. Dorronsoro, F. Ginel, C. Sgñchez, and C. S. Cruz, "Neural fraud detection in credit card operations," *IEEE Transactions on Neural Networks* vol. 8, pp. 827-834, 1997.
- [26] I. C. Yeh and C.-h. Lien, "The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients," *Expert Systems with Applications*, vol. 36, pp. 2473-2480, 2009.
- [27] J. T. S. Quah and M. Sriganesh, "Real-time credit card fraud detection using computational intelligence," *Expert Systems with Applications*, vol. 35, pp. 1721-1732, 2008.
- [28] D. Olszewski, "Fraud detection using self-organizing map visualizing the user profiles," *Knowledge-Based Systems*, vol. 70, pp. 324-334, 2014.
- [29] S. Bhattacharyya, S. Jha, K. Tharakunnel, and J. C. Westland, "Data mining for credit card fraud: A comparative study," *Decision Support Systems*, vol. 50, pp. 602-613, 2011.
- [30] S. Jou and B. Heberton, "Insurance fraud in Taiwan: Reflections on regulatory effort and criminological complexity," *International Journal of the Sociology of Law*, vol. 35, pp. 127-142, 2007.
- [31] S. Viaene, G. Dedene, and R. A. Derrig, "Auto claim fraud detection using Bayesian learning neural networks," *Expert Systems with Applications*, vol. 29, pp. 653-666, 2005.
- [32] P. L. Brockett, R. A. Derrig, L. L. Golden, A. Levine, and M. Alpert, "Fraud Classification Using Principal Component Analysis of RIDITs," *The Journal of Risk and Insurance*, vol. 69, pp. 341-371, 2002.
- [33] L. Bermúdez, J. M. Pérez, M. Ayuso, E. Gómez, and F. J. Vázquez, "A Bayesian dichotomous model with asymmetric link for fraud in insurance," *Insurance: Mathematics and Economics*, vol. 42, pp. 779-786, 2008.
- [34] S. Tennyson and P. Salsas-Forn, "Claims Auditing in Automobile Insurance: Fraud Detection and Deterrence Objectives," *The Journal of Risk and Insurance*, vol. 69, 2002.
- [35] S. Viaene, R. A. Derrig, and G. Dedene, "A case study of applying boosting naive Bayes to claim fraud diagnosis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 16, 2004.
- [36] D. Enke and S. Thawornwong, "The use of data mining and neural networks for forecasting stock market returns," *Expert Systems with Applications*, vol. 29, pp. 927-940, 2005.
- [37] W.-S. Yang and S.-Y. Hwang, "A process-mining framework for the detection of healthcare fraud and abuse," *Expert Systems with Applications*, vol. 31, pp. 56-68, 2006.
- [38] H. He, J. Wang, W. Graco, and S. Hawkins, "Application of neural networks to detection of medical fraud," *Expert System with Applications*, vol. 13, pp. 329-336, 1997.
- [39] P. Wanke, C. P. Barros, and J. R. Faria, "Financial distress drivers in Brazilian banks: A dynamic slacks approach," *European Journal of Operational Research*, vol. 240, pp. 258-268, 2015.
- [40] S.-W. Lin, Y.-R. Shiue, S.-C. Chen, and H.-M. Cheng, "Applying enhanced data mining approaches in predicting bank performance: A case of Taiwanese commercial banks," *Expert Systems with Applications*, vol. 36, pp. 11543-11551, 2009.
- [41] H. Li, C.-J. Li, X.-J. Wu, and J. Sun, "Statistics-based wrapper for feature selection: An implementation on financial distress identification with support vector machine," *Applied Soft Computing*, vol. 19, pp. 57-67, 2014.
- [42] S.-C. Huang, Y.-C. Tang, C.-W. Lee, and M.-J. Chang, "Kernel local Fisher discriminant analysis based manifold-regularized SVM model for financial distress predictions," *Expert Systems with Applications*, vol. 39, 2012.
- [43] J. K. Bae, "Predicting financial distress of the South Korean manufacturing industries," *Expert Systems with Applications*, vol. 39, pp. 9159-9165, 2012.
- [44] W.-S. Chen and Y.-K. Du, "Using neural networks and data mining techniques for the financial distress prediction model," *Expert Systems with Applications*, vol. 36, pp. 4075-4086, 2009.
- [45] R. Geng, I. Bose, and X. Chen, "Prediction of financial distress: An empirical study of listed Chinese companies using data mining," *European Journal of Operational Research*, vol. 241, pp. 236-247, 2015.
- [46] D. Wu, L. Liang, and Z. Yang, "Analyzing the financial distress of Chinese public companies using probabilistic neural networks and multivariate discriminate analysis," *Socio-Economic Planning Sciences*, vol. 42, pp. 206-220, 2008.
- [47] J. Sun and H. Li, "Data mining method for listed companies' financial distress prediction," *Knowledge-Based Systems*, vol. 21, pp. 1-5, 2008.
- [48] M.-Y. Chen, "Predicting corporate financial distress based on integration of decision tree classification and logistic regression," *Expert Systems with Applications*, vol. 38, pp. 11261-11272, 2011.
- [49] S. Y. Kim and A. Upneja, "Predicting restaurant financial distress using decision tree and AdaBoosted decision tree models," *Economic Modelling*, vol. 36, pp. 354-362, 2014.



- [50] H. Li, J. Sun, and B.-L. Sun, "Financial distress prediction based on OR-CBR in the principle of k-nearest neighbors," *Expert Systems with Applications*, vol. 36, pp. 643–659, 2009.
- [51] T. Xiong, S. Wang, A. Mayers, and E. Monga, "Personal bankruptcy prediction by mining credit card data," *Expert Systems with Applications*, vol. 40, pp. 665–676, 2013.
- [52] A. Chaudhuri and K. De, "Fuzzy Support Vector Machine for bankruptcy prediction," *Applied Soft Computing*, vol. 11, pp. 2472–2486, 2011.
- [53] H.-L. Chen, B. Yang, G. Wang, J. Liu, X. Xu, S.-J. Wang, et al., "A novel bankruptcy prediction model based on an adaptive fuzzy k-nearest neighbor method," *Knowledge-Based Systems*, vol. 24, pp. 1348–1359, 2011.
- [54] J. De Andrés, P. Lorca, F. J. de Cos Juez, and F. Sánchez-Lasheras, "Bankruptcy forecasting: A hybrid approach using Fuzzy c-means clustering and Multivariate Adaptive Regression Splines (MARS)," *Expert Systems with Applications*, vol. 38, pp. 1866–1875, 2011.
- [55] P. du Jardin and E. Séverin, "Predicting corporate bankruptcy using a self-organizing map: An empirical study to improve the forecasting horizon of a financial failure model," *Decision Support Systems*, vol. 51, pp. 701–711, 2011.
- [56] D. L. Olson, D. Delen, and Y. Meng, "Comparative analysis of data mining methods for bankruptcy prediction," *Decision Support Systems*, vol. 52, pp. 464–473, 2012.
- [57] T. Van Gestel, B. Baesens, and D. Martens, "From linear to non-linear kernel based classifiers for bankruptcy prediction," *Neurocomputing*, vol. 73, pp. 2955–2970, 2010.
- [58] P. Ravisankar and V. Ravi, "Financial distress prediction in banks using Group Method of Data Handling neural network, counter propagation neural network and fuzzy ARTMAP," *Knowledge-Based Systems*, vol. 23, pp. 823–831, 2010.
- [59] C.-F. Tsai, "Feature selection in bankruptcy prediction," *Knowledge-Based Systems*, vol. 22, pp. 120–127, 2009.
- [60] A. Costea, "Applying Fuzzy Logic and Machine Learning Techniques in Financial Performance Predictions," *Procedia Economics and Finance*, vol. 10, 2014.
- [61] M. Lam, "Neural network techniques for financial performance prediction: integrating fundamental and technical analysis," *Decision Support Systems*, vol. 37, pp. 567–581, 2004.
- [62] S.-H. Chun and S. H. Kim, "Data mining for financial prediction and trading: application to single and multiple markets," *Expert Systems with Applications*, vol. 26, pp. 131–139, 2004.
- [63] A. Berson, S. Smith, and K. Thearling, *Building data mining applications for CRM*: McGraw-Hill, 2000.
- [64] A. Kracklauer, D. Q. Mills, and D. Seifert, "Customer Management as the Origin of Collaborative Customer Relationship Management," *Collaborative Customer Relationship Management*, pp. 3–6, 2004.
- [65] S.-Y. Kim, T.-S. Jung, E.-H. Suh, and H.-S. Hwang, "Customer segmentation and strategy development based on customer lifetime value: A case study," *Expert Systems with Applications*, vol. 31, pp. 101–107, 2006.
- [66] S. H. Han, S. X. Lu, and S. C. H. Leung, "Segmentation of telecom customers based on customer value by decision tree model," *Expert Systems with Applications*, vol. 39, pp. 3964–3973, 2012.
- [67] S. H. Ha, S. M. Bae, and S. C. Park, "Customer's time-variant purchase behavior and corresponding marketing strategies: an online retailer's case," *Computers & Industrial Engineering*, vol. 43, pp. 801–820, 2002.
- [68] F. Fessant, J. François, and F. Clérot, "Characterizing ADSL customer behaviours by network traffic data-mining," *Annales Des Télécommunications*, vol. 62, pp. 350–368, 2007.
- [69] S. M. Bae, S. C. Park, and S. H. Ha, "Fuzzy Web ad selector based on Web usage mining," *Intelligent Systems, IEEE*, vol. 18, pp. 62 – 69, 2003.
- [70] S. V. Verdu, M. O. Garcia, C. Senabre, A. G. Marin, and F. J. G. Franco, "Classification, Filtering, and Identification of Electrical Customer Load Patterns Through the Use of Self-Organizing Maps," *IEEE Transactions on Power Systems*, vol. 21, pp. 1672–1682, 2006.
- [71] S. Hosseini and A. Ziaei Bideh, "A data mining approach for segmentation-based importance-performance analysis (SOM-BPNN-IPA): a new framework for developing customer retention strategies," *Service Business*, vol. 8, pp. 295–312, 2014.
- [72] J. H. Lee and S. C. Park, "Intelligent profitable customers segmentation system based on business intelligence tools," *Expert Systems with Applications*, vol. 29, pp. 145–152, 2005.
- [73] H. Hwang, T. Jung, and E. Suh, "An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry," *Expert Systems with Applications*, vol. 26, pp. 181–188, 2004.
- [74] C. Dennis, D. Marsland, and T. Cockett, "Data mining for shopping centres – customer knowledge-management framework," *Journal of Knowledge Management*, vol. 5, pp. 368–374, 2001.
- [75] J.-T. Wei, M.-C. Lee, H.-K. Chen, and H.-H. Wu, "Customer relationship management in the hairdressing industry: An application of data mining techniques," *Expert Systems with Applications*, vol. 40, pp. 7513–7518, 2013.
- [76] M. Khajvand and M. J. Tarokh, "Estimating customer future value of different customer segments based on adapted RFM model in retail banking context," *Procedia Computer Science*, vol. 3, pp. 1327–1332, 2011.
- [77] M. Khajvand, K. Zolfaghar, S. Ashoori, and S. Alizadeh, "Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study," *Procedia Computer Science*, vol. 3, pp. 57–63, 2011.



- [78] C.-H. Cheng and Y.-S. Chen, "Classifying the segmentation of customer value via RFM model and RS theory," *Expert Systems with Applications*, vol. 36, pp. 4176–4184, 2009.
- [79] S.-H. Liao, C.-M. Chen, and C.-H. Wu, "Mining customer knowledge for product line and brand extension in retailing," *Expert Systems with Applications*, vol. 34, pp. 1763–1776, 2008.
- [80] A. Hızıroglu, "Soft computing applications in customer segmentation: State-of-art review and critique," *Expert Systems with Applications*, vol. 40, pp. 6491–6507, 2013.
- [81] Y. Yang and B. Padmanabhan, "GHIC: A Hierarchical Pattern-Based Clustering Algorithm for Grouping Web Transactions," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, pp. 1300–1304, 2005.
- [82] Z. Wang, L. Tu, Z. Guo, L. T. Yang, and B. Huang, "Analysis of user behaviors by mining large network data sets," *Future Generation Computer Systems*, vol. 37, pp. 429–437, 2014.
- [83] F.-M. Hsu, L.-P. Lu, and C.-M. Lin, "Segmenting customers by transaction data with concept hierarchy," *Expert Systems with Applications*, vol. 39, pp. 6221–6228, 2012.
- [84] R.-S. Wu and P.-H. Chou, "Customer segmentation of multiple category data in e-commerce using a soft-clustering approach," *Electronic Commerce Research and Applications*, vol. 10, pp. 331–341, 2011.
- [85] Y.-L. Chen, C.-L. Hsu, and S.-C. Chou, "Constructing a multi-valued and multi-labeled decision tree," *Expert Systems with Applications*, vol. 25, pp. 199–209, 2003.
- [86] C.-H. Wu, S.-C. Kao, Y.-Y. Su, and C.-C. Wu, "Targeting customers via discovery knowledge for the insurance industry," *Expert Systems with Applications*, vol. 29, pp. 291–299, 2005.
- [87] J. X. Yu, O. Yuming, C. Zhang, and S. Zhang, "Identifying interesting visitors through Web log classification," *IEEE Intelligent Systems*, vol. 20, pp. 55–59, 2005.
- [88] S. C. Lee, Y. H. Suh, J. K. Kim, and K. J. Lee, "A cross-national market segmentation of online game industry using SOM," *Expert Systems with Applications*, vol. 27, 2004.
- [89] Y.-S. Chen, C.-H. Cheng, C.-J. Lai, C.-Y. Hsu, and H.-J. Syu, "Identifying patients in target customer segments using a two-stage clustering-classification approach: A hospital-based assessment," *Computers in Biology and Medicine*, vol. 42, pp. 213–221, 2012.
- [90] G. Mansingh, L. Rao, K.-M. Osei-Bryson, and A. Mills, "Profiling internet banking users: A knowledge discovery in data mining process model based approach," *Information Systems Frontiers*, 2013.
- [91] C.-W. Chang, C.-T. Lin, and L.-Q. Wang, "Mining the text information to optimizing the customer relationship management," *Expert Systems with Applications*, vol. 36, pp. 1433–1443, 2009.
- [92] T. Nakahara and K. Yada, "Analyzing consumers' shopping behavior using RFID data and pattern mining," *Advances in Data Analysis and Classification*, vol. 6, pp. 355–365, 2012.
- [93] C.-K. Huang, T.-Y. Chang, and B. Narayanan, "Mining the change of customer behavior in dynamic markets," *Information Technology and Management*, 2014.
- [94] E. Apeh and B. Gabrys, "Detecting and Visualizing the Change in Classification of Customer Profiles based on Transactional Data," *Evolving Systems*, vol. 4, pp. 27–42, 2013.
- [95] E. Apeh, B. Gabrys, and A. Schierz, "Customer profile classification: To adapt classifiers or to relabel customer profiles?," *Neurocomputing*, vol. 132, pp. 3–13, 2014.
- [96] L. B. Romdhane, N. Fadhel, and B. Ayeb, "An efficient approach for building customer profiles from business data," *Expert Systems with Applications*, vol. 37, pp. 1573–1585, 2010.
- [97] I. Mahdavi, N. Cho, B. Shirazi, and N. Sahebjamnia, "Designing evolving user profile in e-CRM with dynamic clustering of Web documents," *Data & Knowledge Engineering*, vol. 65, pp. 355–372, 2008.
- [98] K.-W. Cheung, J. T. Kwok, M. H. Law, and K.-C. Tsui, "Mining customer product ratings for personalized marketing," *Decision Support Systems*, vol. 35, pp. 231–243, 2003.
- [99] Z. He, X. Xu, J. Z. Huang, and S. Deng, "Mining class outliers: concepts, algorithms and applications in CRM," *Expert Systems with Applications*, vol. 27, pp. 681–697, 2004.
- [100] S.-H. Liao and Y.-J. Chen, "Mining customer knowledge for electronic catalog marketing," *Expert Systems with Applications*, vol. 27, pp. 521–532, 2004.
- [101] A. Prinzie and D. Van den Poel, "Constrained optimization of data-mining problems to improve model performance: A direct-marketing application," *Expert Systems with Applications*, vol. 29, pp. 630–640, 2005.
- [102] B. Baesens, S. Viaene, D. Van den Poel, J. Vanthienen, and G. Dedene, "Bayesian neural network learning for repeat purchase modelling in direct marketing," *European Journal of Operational Research*, vol. 138, pp. 191–211, 2002.
- [103] Y. Kim and W. N. Street, "An intelligent system for customer targeting: a data mining approach," *Decision Support Systems*, vol. 37, pp. 215–228, 2004.
- [104] X. Hu, "A Data Mining Approach for Retailing Bank Customer Attrition Analysis," *Applied Intelligence*, vol. 22, pp. 47–60, 2005.
- [105] W. Buckinx, E. Moons, D. Van den Poel, and G. Wets, "Customer-adapted coupon targeting using feature selection," *Expert Systems with Applications*, vol. 26, pp. 509–518, 2004.
- [106] H. Ahn, K.-j. Kim, and I. Han, "Hybrid genetic algorithms and case-based reasoning systems for customer classification," *Expert Systems with Applications*, vol. 23, pp. 127–144, 2006.



- [107] C. Chiu, "A case-based customer classification approach for direct marketing," *Expert Systems with Applications*, vol. 22, pp. 163–168, 2002.
- [108] L. A. Cox, Jr., "Data Mining and Causal Modeling of Customer Behaviors," *Telecommunication Systems*, vol. 21, pp. 349–381, 2002.
- [109] S. Douglas, D. Agarwal, T. Alonso, R. M. Bell, M. Gilbert, D. F. Swayne, et al., "Mining Customer Care Dialogs for "Daily News"," *IEEE Transactions on Speech and Audio Processing*, vol. 13, pp. 652–660, 2005.
- [110] B. Larivière and D. Van den Poel, "Predicting customer retention and profitability by using random forests and regression forests techniques," *Expert Systems with Applications*, vol. 29, pp. 472–484, 2005.
- [111] Y.-F. Wang, D.-A. Chiang, M.-H. Hsu, C.-J. Lin, and I. L. Lin, "A recommender system to avoid customer churn: A case study," *Expert Systems with Applications*, vol. 36, pp. 8071–8075, 2009.
- [112] B.-H. Chu, M.-S. Tsai, and C.-S. Ho, "Toward a hybrid data mining model for customer retention," *Knowledge-Based Systems*, vol. 20, pp. 703–718, 2007.
- [113] Y. A. Kim, H. S. Song, and S. H. Kim, "Strategies for preventing defection based on the mean time to defection and their implementations on a self-organizing map," *Expert Systems with Applications*, vol. 22, pp. 265–278, 2005.
- [114] H. Song, J. Kim, Y. Cho, and S. Kim, "A Personalized Defection Detection and Prevention Procedure based on the Self-Organizing Map and Association Rule Mining: Applied to Online Game Site," *Artificial Intelligence Review*, vol. 21, pp. 161–184, 2004.
- [115] H. Cassab and D. L. MacLachlan, "Interaction fluency: a customer performance measure of multichannel service," *International Journal of Productivity and Performance Management*, vol. 55, pp. 555–568, 2006.
- [116] W. Buckinx and D. Van den Poel, "Customer base analysis: partial defection of behaviourally loyal clients in a non-contractual FMCG retail setting," *European Journal of Operational Research*, vol. 164, pp. 252–268, 2005.
- [117] P. Datta, B. Masand, D. R. Mani, and B. Li, "Automated Cellular Modeling and Prediction on a Large Scale," *Artificial Intelligence Review*, vol. 14, pp. 485–502, 2000.
- [118] E. Kim, W. Kim, and Y. Lee, "Combination of multiple classifiers for the customer's purchase behavior prediction," *Decision Support Systems*, vol. 34, pp. 167–175, 2003.
- [119] S. M. S. Hosseini, A. Maleki, and M. R. Gholamian, "Cluster analysis using data mining approach to develop CRM methodology to assess the customer loyalty," *Expert Systems with Applications*, vol. 37, pp. 5259–5264, 2010.
- [120] K. Coussement and D. Van den Poel, "Integrating the voice of customers through call center emails into a decision support system for churn prediction," *Information & Management*, vol. 45, pp. 164–174, 2008.
- [121] J. Burez and D. Van den Poel, "Handling class imbalance in customer churn prediction," *Expert Systems with Applications*, vol. 36, pp. 4626–4636, 2009.
- [122] N. Lu, H. Lin, J. Lu, and G. Zhang, "A Customer Churn Prediction Model in Telecom Industry Using Boosting," *Ieee Transactions on Industrial Informatics*, vol. 10, pp. 1659–1665, May 2014.
- [123] B. Larivière and D. Van den Poel, "Investigating the role of product features in preventing customer churn, by using survival analysis and choice modeling: The case of financial services," *Expert Systems with Applications*, vol. 27, pp. 277–285, 2004.
- [124] A. Lemmens and C. Croux, "Bagging and Boosting Classification Trees to Predict Churn," *Journal of Marketing Research*, vol. 43, pp. 276–286, 2006.
- [125] K. Coussement and D. Van den Poel, "Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques," *Expert Systems with Applications*, vol. 34, pp. 313–327, 2008.
- [126] S. A. Neslin, S. Gupta, W. Kamakura, J. Lu, and C. H. Mason, "Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models," *Journal of Marketing Research* vol. 43, pp. 204–211, 2006.
- [127] M. C. Mozer, R. Wolniewicz, D. B. Grimes, E. Johnson, and H. Kaushansky, "Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry," *IEEE Transactions on Neural Networks*, vol. 11, pp. 690–696, 2000.
- [128] C.-P. Wei and I. T. Chiu, "Turning telecommunications call details to churn prediction: a data mining approach," *Expert Systems with Applications*, vol. 23, pp. 103–112, 2002.
- [129] S.-Y. Hung, D. C. Yen, and H.-Y. Wang, "Applying data mining to telecom churn management," *Expert Systems with Applications*, vol. 31, pp. 515–524, 2006.
- [130] C.-F. Tsai and Y.-H. Lu, "Customer churn prediction by hybrid neural networks," *Expert Systems with Applications*, vol. 36, pp. 12547–12553, 2009.
- [131] M.-C. Chen, A.-L. Chiu, and H.-H. Chang, "Mining changes in customer behavior in retail marketing," *Expert Systems with Applications*, vol. 28, pp. 773–781, 2005.
- [132] G. Adomavicius and A. Tuzhilin, "Expert-Driven Validation of Rule-Based User Models in Personalization Applications," *Data Mining and Knowledge Discovery*, vol. 5, pp. 33–58, 2001.
- [133] W.-H. Au and C. C. C. Chan, "Mining fuzzy association rules in a bank-account database," *IEEE Transactions on Fuzzy Systems*, vol. 11, pp. 238 – 248, 2003.



- [134] A. Demiriz, "Enhancing Product Recommender Systems on Sparse Binary Data," *Data Mining and Knowledge Discovery*, vol. 9, pp. 147-170, 2004.
- [135] J. Jiao, Y. Zhang, and M. Helander, "A Kansei mining system for affective design," *Expert Systems with Applications*, vol. 30, pp. 658-673, 2006.
- [136] C. H. Lee, Y. H. Kim, and P. K. Rhee, "Web personalization expert with combining collaborative filtering and association rule mining technique," *Expert Systems with Applications*, vol. 21, pp. 131-137, 2001.
- [137] Y.-F. Wang, Y.-L. Chuang, M.-H. Hsu, and H.-C. Keh, "A personalized recommender system for the cosmetic business," *Expert Systems with Applications*, vol. 26, pp. 427-434, 2004.
- [138] S.-H. Liao, C.-M. Chen, C.-L. Hsieh, and S.-C. Hsiao, "Mining information users' knowledge for one-to-one marketing on information appliance," *Expert Systems with Applications*, vol. 36, pp. 4967-4979, 2009.
- [139] H. S. Song, J. k. Kim, and S. H. Kim, "Mining the change of customer behavior in an internet shopping mall," *Expert Systems with Applications*, vol. 21, pp. 157-168, 2001.
- [140] I. S. Y. Kwan, J. Fong, and H. K. Wong, "An e-customer behavior model with online analytical mining for internet marketing planning," *Decision Support Systems*, vol. 41, pp. 189-204, 2005.
- [141] Y. B. Cho, Y. H. Cho, and S. H. Kim, "Mining changes in customer buying behavior for collaborative recommendations," *Expert Systems with Applications*, vol. 28, pp. 359-369, 2005.
- [142] N.-C. Hsieh, "An integrated data mining and behavioral scoring model for analyzing bank customers," *Expert Systems with Applications*, vol. 27, pp. 623-633, 2004.
- [143] J. H. Lee and S. C. Park, "Agent and data mining based decision support system and its adaptation to a new customer-centric electronic commerce," *Expert Systems with Applications*, vol. 25, pp. 619-635, 2003.
- [144] S.-H. Min and I. Han, "Detection of the customer time-variant pattern for improving recommender systems.," *Expert Systems with Applications*, vol. 28, pp. 189-199, 2005.
- [145] S. H. Ha, "Digital content recommender on the Internet," *IEEE Intelligent Systems*, vol. 21, pp. 70 - 77, 2006.
- [146] J. K. Kim, H. S. Song, T. S. Kim, and H. K. Kim, "Detecting the change of customer behavior based on decision tree analysis," *Expert Systems*, vol. 22, pp. 193-205, 2005.
- [147] S. E. Chang, S. W. Changchien, and R.-H. Huang, "Assessing users' productspecific knowledge for personalization in electronic commerce," *Expert Systems with Applications*, vol. 30, pp. 682-693, 2006.
- [148] R. J. Kuo, J. L. Liao, and C. Tu, "Integration of ART2 neural network and genetic K-means algorithm for analyzing Web browsing paths in electronic commerce," *Decision Support Systems*, vol. 40, pp. 355-374, 2005.
- [149] Y. H. Cho and J. K. Kim, "Application of Web usage mining and product taxonomy to collaborative recommendations in e-commerce," *Expert Systems with Applications*, vol. 26, pp. 233-246, 2004.
- [150] S. M. Bae, S. H. Ha, and S. C. Park, "A web-based system for analyzing the voices of call center customers in the service industry," *Expert Systems with Applications*, vol. 28, pp. 29-41, 2005.
- [151] B. Shim, K. Choi, and Y. Suh, "CRM strategies for a small-sized online shopping mall based on association rules and sequential patterns," *Expert Systems with Applications*, vol. 39, pp. 7736-7742, 2012.
- [152] K. Ng and H. Liu, "Customer Retention via Data Mining," *Artificial Intelligence Review*, vol. 14, pp. 569-590, 2000.
- [153] Z.-Y. Chen, Z.-P. Fan, and M. Sun, "A hierarchical multiple kernel support vector machine for customer churn prediction using longitudinal behavioral data," *European Journal of Operational Research*, vol. 223, pp. 461-472, 2012.
- [154] J. H. Drew, D. R. Mani, A. L. Betz, and P. Datta, "Targeting Customers with Statistical and Data-Mining Techniques," *Journal of Service Research*, vol. 3, pp. 205-219, 2001.
- [155] C. J. Cheng, S. W. Chiu, C. B. Cheng, and J. Y. Wu, "Customer lifetime value prediction by a Markov chain based data mining model: Application to an auto repair and maintenance company in Taiwan," *Scientia Iranica*, vol. 19, pp. 849-855, 2012.
- [156] Z.-Y. Chen and Z.-P. Fan, "Dynamic customer lifetime value prediction using longitudinal data: An improved multiple kernel SVR approach," *Knowledge-Based Systems*, vol. 43, pp. 123-134, 2013.
- [157] D.-R. Liu and Y.-Y. Shih, "Integrating AHP and data mining for product recommendation based on customer lifetime value," *Information & Management*, vol. 42, pp. 387-400, 2005.
- [158] P. C. Verhoef and B. Donkers, "Predicting customer potential value an application in the insurance industry.," *Decision Support Systems*, vol. 32, pp. 189-199, 2001.
- [159] B. Baesens, G. Verstraeten, D. Van den Poel, M. Egmont-Petersen, P. Van Kenhove, and J. Vanthienen, "Bayesian network classifiers for identifying the slope of the customer lifecycle of long-life customers," *European Journal of Operational Research*, vol. 156, pp. 508-523, 2004.
- [160] Y.-H. Liang, "Integration of data mining technologies to analyze customer value for the automotive maintenance industry," *Expert Systems with Applications*, vol. 37, pp. 7489-7496, 2010.
- [161] W.-Y. Chiang, "To mine association rules of customer values via a data mining procedure with improved model: An empirical case study," *Expert Systems with Applications*, vol. 38, pp. 1716-1722, 2011.
- [162] K. Wong, S. Zhou, Q. Yang, and J. Yeung, "Mining Customer Value: From Association Rules to Direct Marketing," *Data Mining and Knowledge Discovery*, vol. 11, pp. 57-79, 2005.



- [163] R.-W. Wong, A.-C. Fu, and K. Wang, "Data Mining for Inventory Item Selection with Cross-Selling Considerations," *Data Mining and Knowledge Discovery*, vol. 11, pp. 81-112, 2005.
- [164] S. W. Changchien, C.-F. Lee, and Y.-J. Hsu, "On-line personalized sales promotion in electronic commerce," *Expert Systems with Applications*, vol. 27, pp. 35-52, 2004.
- [165] S.-h. Liao, Y.-J. Chen, and M.-y. Deng, "Mining customer knowledge for tourism new product development and customer relationship management," *Expert Systems with Applications*, vol. 37, pp. 4212-4223, 2010.
- [166] C. C. Aggarwal, C. Procopiuc, and P. S. Yu, "Finding localized associations in market basket data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 14, pp. 51-62, 2002.
- [167] T. Brijs, G. Swinnen, K. Vanhoof, and G. Wets, "Building an Association Rules Framework to Improve Product Assortment Decisions," *Data Mining and Knowledge Discovery*, vol. 8, pp. 7-23, 2004.
- [168] N. Jukić and S. Nestorov, "Comprehensive data warehouse exploration with qualified association-rule mining," *Decision Support Systems*, vol. 42, pp. 859-878, 2006.
- [169] P. Giudici and G. Passerone, "Data mining of association structures to model consumer behaviour," *Computational Statistics & Data Analysis*, vol. 38, pp. 533-541, 2002.
- [170] Y.-L. Chen, K. Tang, R.-J. Shen, and Y.-H. Hu, "Market basket analysis in a multiple store environment," *Decision Support Systems*, vol. 40, pp. 339-354, 2005.
- [171] A. Kusiak and M. Smith, "Data mining in design of products and production systems," *Annual Reviews in Control*, vol. 31, pp. 147-156, 2007.
- [172] P. Kim and Y. Ding, "Optimal Engineering System Design Guided by Data-Mining Methods," *Technometrics*, vol. 47, pp. 336-348, 2005.
- [173] O. Torkul, I. H. Cedimoglu, and A. K. Geyik, "An application of fuzzy clustering to manufacturing cell design," *Journal of Intelligent and Fuzzy Systems*, vol. 17, pp. 173-181, 2006.
- [174] O. G. Ali and C. Yu-To, "Design quality and robustness with neural networks," *IEEE Transactions on Neural Networks*, vol. 10, pp. 1518 - 1527, 1999.
- [175] J.-H. Park and K.-K. Seo, "A knowledge-based approximate life cycle assessment system for evaluating environmental impacts of product design alternatives in a collaborative design environment," *Advanced Engineering Informatics*, vol. 20, pp. 147-154, 2006.
- [176] C.-Y. Tsai and C. A. Chang, "Fuzzy neural networks for intelligent design retrieval using associative manufacturing features," *Journal of Intelligent Manufacturing*, vol. 14, pp. 183-195, 2003.
- [177] J. K. Bae and J. Kim, "Product development with data mining techniques: A case on design of digital camera," *Expert Systems with Applications*, vol. 38, 2011.
- [178] C. J. Romanowski and R. Nagi, "On comparing bills of materials: a similarity/distance measure for unordered trees," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 35, pp. 249 - 260, 2005.
- [179] B. Agard and A. Kusiak *, "A Data Mining based Methodology for the Design of Product Families," *International Journal of Production Research*, vol. 42, pp. 2955-2969, 2004.
- [180] Y. Jin and Y. Ishino, "DAKA: design activity knowledge acquisition through data-mining," *International Journal of Production Research*, vol. 44, pp. 2813-2837, 2006.
- [181] Y. Liu, W. Feng Lu, and H. Tong Loh, "A framework of information and knowledge management for product design and development — A text mining approach," *Information Control Problems in Manufacturing*, pp. 635-640, 2006.
- [182] N. Morar, R. Roy, J. Mehnen, L. E. Redding, and A. Harrison, "Data Mining and Knowledge Reuse for the Initial Systems Design and Manufacturing: Aero-engine Service Risk Drivers," *Procedia CIRP*, vol. 11, pp. 130-134, 2013.
- [183] M.-C. Chen, "Configuration of cellular manufacturing systems using association rule induction," *International Journal of Production Research*, vol. 41, pp. 381-395, 2003.
- [184] C. Da Cunha, B. Agard, and A. Kusiak, "Data mining for improvement of product quality," *International Journal of Production Research*, vol. 44, pp. 4027-4041, 2006.
- [185] J. Yu, L. Xi, and X. Zhou, "Intelligent monitoring and diagnosis of manufacturing processes using an integrated approach of KBANN and GA," *Computers in Industry*, vol. 59, pp. 489-501, 2008.
- [186] J. Sun, G. S. Hong *, M. Rahman, and Y. S. Wong, "Improved performance evaluation of tool condition identification by manufacturing loss consideration," *International Journal of Production Research*, vol. 43, pp. 1185-1204, 2005.
- [187] J. H. Lee, S. J. Yu, and S. C. Park, "Design of intelligent data sampling methodology based on data mining," *IEEE Transactions on Robotics and Automation*, vol. 17, pp. 637-649, 2001.
- [188] C.-X. J. Feng and X.-F. Wang, "Data mining techniques applied to predictive modeling of the knurling process," *IEEE Transactions*, vol. 36, pp. 253-263, 2004.
- [189] Z. J. Pasek, "Exploration of Rough Sets Theory Use for Manufacturing Process Monitoring," *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, vol. 220, pp. 365-374, 2006.



- [190] I. Abu-Mahfouz and A. Banerjee, "Drill Wear Feature Identification under Varying Cutting Conditions Using Vibration and Cutting Force Signals and Data Mining Techniques," *Procedia Computer Science*, vol. 36, pp. 556–563, 2014.
- [191] W.-C. Chen, S.-S. Tseng, and C.-Y. Wang, "A novel manufacturing defect detection method using association rule mining techniques," *Expert Systems with Applications*, vol. 29, pp. 807–815, 2005.
- [192] C.-H. Wang, W. Kuo, and H. Bensmail, "Detection and classification of defect patterns on semiconductor wafers," 38, vol. 12, 2006.
- [193] T. W. Liao, D. M. Li, and Y. M. Li, "Detection of welding flaws from radiographic images with fuzzy clustering methods," *Fuzzy Sets and Systems*, vol. 108, pp. 145–158, 1999.
- [194] D. A. Koonce and S. C. Tsai, "Using data mining to find patterns in genetic algorithm solutions to a job shop schedule," *Computers & Industrial Engineering*, vol. 38, pp. 361–374, 2000.
- [195] C. Kwak and Y. Yih, "Data-mining approach to production control in the computer-integrated testing cell," *IEEE Transactions on Robotics and Automation*, vol. 20, pp. 107–116, 2004.
- [196] A. Öztürk, S. Kayaligil, and N. E. Özdemirel, "Manufacturing lead time estimation using data mining," *European Journal of Operational Research*, vol. 173, pp. 683–700, 2006.
- [197] K. J. Wang, J. C. Chen, and Y. S. Lin, "A Hybrid Knowledge Discovery Model Using Decision Tree and Neural Network for Selecting Dispatching Rules of A Semiconductor Final Testing Factory," *Production Planning and Control*, vol. 16, 2005.
- [198] S. Chen, G. Xiaohong, Z. Qianchuan, and Y. C. Ho, "Machine learning approach for determining feasible plans of a remanufacturing system," *IEEE Transactions on Automation Science and Engineering*, vol. 2, pp. 262–275, 2005.
- [199] M. K. Jeong, J.-C. Lu, X. Huo, B. Vidakovic, and C. Di, "Wavelet-Based Data Reduction Techniques for Process Fault Detection," *Technometrics*, vol. 48, pp. 26–40, 2006.
- [200] K. Irani, J. Cheng, U. M. Fayyad, and Z. Qian, "Applying machine learning to semiconductor manufacturing," *IEEE Expert*, vol. 8, pp. 41–47, 1993.
- [201] V. A. Skormin, V. I. Gorodetski, and L. J. Popyack, "Data mining technology for failure prognostic of avionics," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, pp. 388–403, 2002.
- [202] A. Rojas and A. K. Nandi, "Practical scheme for fast detection and classification of rolling-element bearing faults using support vector machines," *Mechanical Systems and Signal Processing*, vol. 20, pp. 1523–1536, 2006.
- [203] T.-S. Li and C.-L. Huang, "Defect spatial pattern recognition using a hybrid SOM–SVM approach in semiconductor manufacturing," *Expert Systems with Applications*, vol. 36, pp. 374–385, 2009.
- [204] Z. Geng and Q. Zhu, "Rough set-based heuristic hybrid recognizer and its application in fault diagnosis," *Expert Systems with Applications*, vol. 36, pp. 2711–2718, 2009.
- [205] M. Demetgul, K. Yildiz, S. Taskin, I. N. Tansel, and O. Yazicioglu, "Fault diagnosis on material handling system using feature selection and data mining techniques," *Measurement*, vol. 55, pp. 15–24, 2014.
- [206] T.-L. Chiang, C.-T. Su, T.-S. Li, and R. C. C. Huang, "Improvement of Process Capability Through Neural Networks and Robust Design: A Case Study," *Quality Engineering*, vol. 14, pp. 313–318, 2002.
- [207] B. Kamsu-Foguem, F. Rigal, and F. Mauget, "Mining association rules for the quality improvement of the production process," *Expert Systems with Applications*, vol. 40, pp. 1034–1045, 2013.
- [208] T.-N. Tsai, "Development of a soldering quality classifier system using a hybrid data mining approach," *Expert Systems with Applications*, vol. 39, pp. 5727–5738, 2012.
- [209] Y. M. Sebzalli and X. Z. Wang, "Knowledge discovery from process operational data using PCA and fuzzy clustering," *Engineering Applications of Artificial Intelligence*, vol. 14, pp. 607–616, 2001.
- [210] W.-C. Chen, A. H. I. Lee, W.-J. Deng, and K.-Y. Liu, "The implementation of neural network for semiconductor PECVD process," *Expert Systems with Applications*, vol. 32, pp. 1148–1153, 2007.
- [211] D. Lieber, M. Stolpe, B. Konrad, J. Deuse, and K. Morik, "Quality Prediction in Interlinked Manufacturing Processes based on Supervised & Unsupervised Machine Learning," *Procedia CIRP*, vol. 7, pp. 193–198, 2013.
- [212] S. Ferreira, B. Sierra, I. Irigoien, and E. Gorritxategi, "Data mining for quality control: Burr detection in the drilling process," *Computers & Industrial Engineering*, vol. 60, pp. 801–810, 2011.
- [213] D. S. Chang and S.-T. Jiang, "Assessing quality performance based on the on-line sensor measurements using neural networks," *Computers & Industrial Engineering*, vol. 42, pp. 417–424, 2002.
- [214] C.-F. Chien, W.-C. Wang, and J.-C. Cheng, "Data mining for yield enhancement in semiconductor manufacturing and an empirical study," *Expert Systems with Applications*, vol. 33, pp. 192–198, 2007.
- [215] C.-f. Chien, H.-c. Li, and A. Jeang, "Data mining for improving the solder bumping process in the semiconductor packaging industry," vol. 14, pp. 43–57, 2006.
- [216] R. M. Dabbas and H.-N. Chen, "Mining semiconductor manufacturing data for productivity improvement — an integrated relational database approach," *Computers in Industry*, vol. 45, pp. 29–44, 2001.
- [217] B.-S. Kang, J.-H. Lee, C.-K. Shin, S.-J. Yu, and S.-C. Park, "Hybrid machine learning system for integrated yield management in semiconductor manufacturing," *Expert Systems with Applications*, vol. 15, pp. 123–132, 1998.



- [218] D. Braha and A. Shmilovici, "Data mining for improving a cleaning process in the semiconductor industry," *IEEE Transactions on Semiconductor Manufacturing*, vol. 15, pp. 91-101, 2002.
- [219] C.-W. Liu and C.-F. Chien, "An intelligent system for wafer bin map defect diagnosis: An empirical study for semiconductor manufacturing," *Engineering Applications of Artificial Intelligence*, vol. 26, pp. 1479-1486, 2013.
- [220] M. Last and A. Kandel, "Discovering useful and understandable patterns in manufacturing data," *Robotics and Autonomous Systems*, vol. 49, pp. 137-152, 2004.
- [221] K.-L. Hsieh, "Incorporating ANNs and statistical techniques into achieving process analysis in TFT-LCD manufacturing industry," *Robotics and Computer-Integrated Manufacturing*, vol. 26, pp. 92-99, 2010.
- [222] S.-C. Hsu and C.-F. Chien, "Hybrid data mining approach for pattern extraction from wafer bin map to improve yield in semiconductor manufacturing," *International Journal of Production Economics*, vol. 107, pp. 88-103, 2007.
- [223] C. J. McDonald, "New tools for yield improvement in integrated circuit manufacturing: can they be applied to reliability?," *Microelectronics Reliability*, vol. 39, pp. 731-739, 1999.
- [224] T.-S. Li, C.-L. Huang, and Z.-Y. Wu, "Data mining using genetic programming for construction of a semiconductor manufacturing yield rate prediction system," *Journal of Intelligent Manufacturing*, vol. 17, pp. 355-361, 2006.
- [225] M. Gardner and J. Bieker, "Data mining solves tough semiconductor manufacturing problems," presented at the Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, Boston, Massachusetts, USA, 2000.
- [226] D.-H. Baek, I.-J. Jeong, and C. H. Han, "Application of Data Mining for Improving Yield in Wafer Fabrication System," *Computational Science and Its Applications - ICCSA*, vol. 3483, pp. 222-231, 2005.
- [227] C. Zhou, P. Nelson, W. Xiao, T. M. Tirpak, and S. A. Lane, "An intelligent data mining system for drop test analysis of electronic products," *IEEE Transactions on Electronics Packaging Manufacturing*, vol. 24, pp. 222-231, 2001.
- [228] C. Çiflikli and E. Kahya-Özyirmidokuz, "Implementing a data mining solution for enhancing carpet manufacturing productivity," *Knowledge-Based Systems*, vol. 23, pp. 783-788, 2010.
- [229] L. Varga, K. Czinege, and I. Ieee, "Electricity consumer characterisation in liberalised market based on data mining techniques," 2007 42nd International Universities Power Engineering Conference, Vols 1-3, pp. 78-82, 2007.
- [230] Y.-L. Chen, T. C.-K. Huang, and S.-K. Chang, "A novel approach for discovering retail knowledge with price information from transaction databases," *Expert Systems with Applications*, vol. 34, pp. 2350-2359, 2008.
- [231] K.-Y. Wang, I. H. Ting, and H.-J. Wu, "Discovering interest groups for marketing in virtual communities: An integrated approach," *Journal of Business Research*, vol. 66, pp. 1360-1366, 2013.
- [232] Y. Ni and F. Fan, "A two-stage dynamic sales forecasting model for the fashion retail," *Expert Systems with Applications*, vol. 38, pp. 1529-1536, 2011.
- [233] Z. X. Guo, W. K. Wong, and M. Li, "A multivariate intelligent decision-making model for retail sales forecasting," *Decision Support Systems*, vol. 55, pp. 247-255, 2013.
- [234] D. Bhanu and S. P. Madheswari, "Retail Marketing Segmentation and Customer Profiling for Forecasting Sales," *Recent Advances in Networking, Vlsi and Signal Processing*, pp. 230-239, 2010.
- [235] A. Demiriz, "Demand Forecasting based on Pairwise Item Associations," *Procedia Computer Science*, vol. 36, pp. 261-268, 2014.
- [236] D. Papakiriakopoulos, "Predict on-shelf product availability in grocery retailing with classification methods," *Expert Systems with Applications*, vol. 39, pp. 4473-4482, 2012.
- [237] L.-D. Chen, T. Sakaguchi, and M. N. Frolick, "Data Mining Methods, Applications, and Tools," *Information Systems Management*, vol. 17, pp. 65-70, 2000.
- [238] S.-H. Liao, W.-J. Chang, and C.-C. Lee, "Mining marketing maps for business alliances," *Expert Systems with Applications*, vol. 35, pp. 1338-1350, 2008.
- [239] S.-h. Liao, Y.-j. Chen, and H.-h. Hsieh, "Mining customer knowledge for direct selling and marketing," *Expert Systems with Applications*, vol. 38, pp. 6059-6069, 2011.
- [240] S.-h. Liao, Y.-j. Chen, and Y.-t. Lin, "Mining customer knowledge to implement online shopping and home delivery for hypermarkets," *Expert Systems with Applications*, vol. 38, pp. 3982-3991, 2011.
- [241] S.-H. Liao, J.-L. Chen, and T.-Y. Hsu, "Ontology-based data mining approach implemented for sport marketing," *Expert Systems with Applications*, vol. 36, pp. 11045-11056, 2009.
- [242] D. M. Kyun, S. Y. Sohn, and J. S. Kim, "Multi-level association rules for MP3P marketing strategies based on extensive marketing survey data," *Expert Systems with Applications*, vol. 36, pp. 11582-11588, 2009.
- [243] C.-C. Yu and C.-S. Wang, "A hybrid mining approach for optimizing returns policies in e-retailing," *Expert Systems with Applications*, vol. 35, pp. 1575-1582, 2008.
- [244] M. Kargari and M. M. Sepehri, "Stores clustering using a data mining approach for distributing automotive spare-parts to reduce transportation costs," *Expert Systems with Applications*, vol. 39, pp. 4740-4748, 2012.
- [245] I. C. Yeh, C.-h. Lien, T.-M. Ting, and C.-H. Liu, "Applications of web mining for marketing of online bookstores," *Expert Systems with Applications*, vol. 36, pp. 11249-11256, 2009.



- [246] C.-C. Hsu and Y.-C. Chen, "Mining of mixed data with application to catalog marketing," *Expert Systems with Applications*, vol. 32, pp. 12–23, 2007.
- [247] R. F. Chang and C. N. Lu, "Load profile assignment of low voltage customers for power retail market applications," *IEEE Proceedings-Generation, Transmission and Distribution*, vol. 150, pp. 263–267, 2003.
- [248] Y. Boztuğ and T. Reutterer, "A combined approach for segment-specific market basket analysis," *European Journal of Operational Research*, vol. 187, pp. 294–312, 2008.
- [249] A. Seret, T. Verbraken, S. Versailles, and B. Baesens, "A new SOM-based method for profile generation: Theory and an application in direct marketing," *European Journal of Operational Research*, vol. 220, pp. 199–209, 2012.
- [250] Y. A. Kim, H. S. Song, and S. H. Kim, "A new marketing strategy map for direct marketing," *Knowledge-Based Systems*, vol. 22, pp. 327–335, 2009.
- [251] D. Lee, S.-H. Park, and S. Moon, "Utility-based association rule mining: A marketing solution for cross-selling," *Expert Systems with Applications*, vol. 40, pp. 2715–2725, 2013.
- [252] A. Mild and T. Reutterer, "An improved collaborative filtering approach for predicting cross-category purchases based on binary market basket data," *Journal of Retailing and Consumer Services*, vol. 10, pp. 123–133, 2003.
- [253] H. Tang, S. S. Liao, and S. X. Sun, "A prediction framework based on contextual data to support Mobile Personalized Marketing," *Decision Support Systems*, vol. 56, pp. 234–246, 2013.
- [254] Y.-L. Chen, J.-M. Chen, and C.-W. Tung, "A data mining approach for retail knowledge discovery with consideration of the effect of shelf-space adjacency on sales," *Decision Support Systems*, vol. 42, pp. 1503–1520, 2006.
- [255] K.-I. Ahn, "Effective product assignment based on association rule mining in retail," *Expert Systems with Applications*, vol. 39, pp. 12551–12556, 2012.
- [256] M.-C. Chen and C.-P. Lin, "A data mining approach to product assortment and shelf space allocation," *Expert Systems with Applications*, vol. 32, pp. 976–986, 2007.
- [257] T. Brijs, G. Swinnen, K. Vanhoof, and G. Wets, "Using association rules for product assortment decisions: a case study," presented at the Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining, San Diego, California, USA, 1999.
- [258] G. Aloysius and D. Binu, "An approach to products placement in supermarkets using PrefixSpan algorithm," *Journal of King Saud University - Computer and Information Sciences*, vol. 25, pp. 77–87, 2013.
- [259] I. F. Videla-Cavieles and S. A. Ríos, "Extending market basket analysis with graph mining techniques: A real case," *Expert Systems with Applications*, vol. 41, pp. 1928–1936, 2014.
- [260] I. Cil, "Consumption universes based supermarket layout through association rule mining and multidimensional scaling," *Expert Systems with Applications*, vol. 39, pp. 8611–8625, 2012.