Financial Profiling for Detecting Operational Risk by Data Mining

Ali Serhan Koyuncugil, PH.D.
Capital Markets Board of Turkey, Turkey

Nermin Ozgulbas, PH.D.
Baskent University, Turkey

Abstract
Basel II Capital Accord has been the center of attention of credit companies, banks and firms. Under Basel II banks will from now on not only consider financial risk of firms but also operational risk thereof before they grant loans to firms. Nowadays, measuring and detecting operational risk is a complicated task fundamental element of business success and as well as hedging financial risk. The aim of this study is profiling the firms according to operational risk factors affected financial performance and detecting the operational risk factors. For this purpose we used CHAID (Chi-Square Automatic Interaction Detector) decision trees–one of the data mining algorithms -, which is one of the best ways to identify financial profiles of firms and determine operational risk factors. The study covered 6185 firms in Organized Industrial Region of Ankara in Turkey. Firms were categorized into 35 profiles on basis of operational risk factors by means of CHAID decision tree algorithm. It was found that firms should emphasize the educational background of managers, status of managers, annual turnover, operating length of firms, makers of financial strategies, expenditure of energy, knowledge about BASEL-II and quality standards, and usage of credit as operational risk factors for hedging operational risk and raising financial performance.

Key Words: Financial Performance, Risk, Operational Risk, Financial Risk, Data Mining, CHAID.

1. Introduction

Risk management has become a vital topic for all institutions. Strategically, asset/liability management systems are important tools for controlling a firm's financial risks. Benefits of risk management can summarize as early warning to avoid distress, road maps for good credit rating, better business decision making, and greater likelihood of achieving business plan and objectives. Nowadays, Basel II Capital Accord, which will become effective in soon, has been the center of attention of credit companies, banks and firms. Basel II is the second of the Basel Accords, which are recommendations on banking laws and regulations issued by the Basel Committee on Banking Supervision. The purpose of Basel II, which was initially published in June 2004, is to create an international standard that banking regulators can use when creating regulations about how much capital banks need to put aside to guard against the types of risks banks face. Basel-II, which introduces risk-based capital management and risk-based credit pricing, would negatively/positively affect amount/price of loans to be utilized by firms. Under Basel-II banks will from now on not only consider financial risk of firms but also operational risk thereof before they grant loans to firms. With minimum financial and operational risk firms would get higher ratings from banks and independent auditing institutions thereby increasing their chances to receive loan facilities with more favorable conditions and minimum costs.

The Basel Committee on Banking Supervision (2003) wants to enhance operational risk assessment efforts by encouraging the industry to develop methodologies and collect data related to managing operational risk. Although operational risk is not a new risk, deregulation and globalization of financial services, together with the growing sophistication of financial technology, new business activities and delivery channels, are making institutions' operational risk profiles more complex.

Nowadays, measuring and detecting operational risk is a complicated task fundamental element of business success and as well as hedging financial risk. The objective of this study is profiling the firms according to operational risk factors affected financial performance and detecting the operational risk factors. For this purpose we used CHAID (Chi-Square Automatic Interaction Detector) decision trees – one of the data mining algorithms -, which is one of the best ways to identify financial profiles of firms and determine operational risk factors. Remaining of this paper is organized as follows: Section 2 presents definition of operational risk. Section 3 contains definition of the data mining and CHAID. Implementation and methodology is presented in Section 4. Section 5 provides the results of the study. Concluding remarks and strategies were suggested in the Conclusion Section.

2. Operational Risk

The Basel Committee on Banking Supervision (2001) has adopted a common industry definition of operational risk, namely: the risk of direct or indirect loss resulting from inadequate or failed internal processes, people and systems or from external events. The definition includes legal risk, which is the risk of loss resulting from failure to comply with laws as well as prudent ethical standards and contractual obligations. It also includes the exposure to litigation from all aspects of an institution's activities. The definition does not include strategic or reputational risks.

Because of the complexity of event which generate operational loss and the heterogeneity of causes, the Committee purposes three approaches which can be, generally, clustered in two main strategies: top–down and bottom –up (Cornalba and Giudicib, 2004):

1. Top–down methods: operational risks are measured and covered at a central level, so local business units (e.g. bank branches) are not involved in the measurement and allocation process. The calculation of the capital requirement is performed using variables that are strongly correlated with risk exposure.
2. Bottom–up methods: differently from the preceding methodologies, operational risk exposures and losses have been broken into a series of standardized 3 business units (called business lines) and into a group of operational risk losses according to the nature of the underlying operational risk event (called event type); operational risks are measured at the level of each business line and then aggregated. Although capital coverage is decided centrally, the contribution of each business line is visible and can be monitored; at the same time it's more expensive to implement, but it allows much better management control and planning in a particular business line.

Although The Basel Committee on Banking Supervision (2001) proposed The Basic Indicator Approach (BIA), The Standardized Approach (TSA), and The Advanced Measurement Approach (AMA) for calculation operational risk, qualitative approaches as self assessment and statistical models can apply by banks and firms. In this study we use data mining for measuring and detecting operational risk.

3. Data Mining and CHAID Decision Trees

The explosive growth in data and database results in the need to develop new technologies and tools to process data into useful information and knowledge intelligently and automatically. Data mining, therefore, has become a research area with increasing importance (Weiss and Indurkhya, 1998; Fayyad *et al.*, 1996; Fayyad *et al.*, 1996; Piatetsky-Shapiro and Frawley, 1991). Kleissner (1998) specified that some business trends like data explosion, business reengineering and organizational decentralization, faster product cycles and globalization and enterprise topologies have made the usage of data mining tools and services mandatory for companies.

Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations (Thearling, 2009).

Nowadays main challenge in the banking and business world is therefore the implementation of risk management systems in order to identify, measure, and control business exposure. Data mining derives its name from the similarities between searching for valuable business information in a large database.

Today, data mining technology integrated measurement of different kinds of is moving into focus to measure and hedging risk. Data mining techniques have been successfully applied like fraud detection and bankruptcy prediction by Tam and Kiang (1992), Lee *et al.* (1996), Kumar *et al*(1997), strategic decision-making by Nazem and Shin (1999) and financial performance by Eklund *et al.* (2003), Hoppszallern (2003), Derby (2003), Chang *et al.* (2003), Kloptchenko *et a.,* (2004), Magnusson *et al.* (2005). Also, some earlier studies of Koyuncugil and Ozgulbas (2006a, 2006b, 2006c, 2007, 2008a, 2008b, 2009) Ozgulbas and Koyuncugil (2006, 2009) conducted on financial performance, financial risk and operational risk of Small and Medium Enterprises (SMEs) and hospitals by data mining.

Fayyad *et al.* (1996), proposed main steps of DM:

- Retrieving the data from a large database.
- Selecting the relevant subset to work with.
- Deciding on the appropriate sampling system, cleaning the data and dealing with missing fields and records.
- Applying the appropriate transformations, dimensionality reduction, and projections.
- Fitting models to the preprocessed data.

Data mining techniques can yield the benefits of automation on existing software and hardware platforms, and can be implemented on new systems as existing platforms are upgraded and new products developed. When data mining tools are implemented on high performance parallel processing systems, they can analyze massive databases in minutes. The most commonly used techniques in data mining are (Thearling, 2009, Koyuncugil, 2006):

- **Artificial neural networks**: Non-linear predictive models that learn through training and resemble biological neural networks in structure.
- **Decision trees**: Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID).
- **Genetic algorithms**: Optimization techniques that use process such as genetic combination, mutation, and natural selection in a design based on the concepts of evolution.
- **Nearest neighbor method**: A technique that classifies each record in a dataset based on a combination of the classes of the k record(s) most similar to it in a historical dataset. Sometimes called the k-nearest neighbor technique.
- **Rule induction**: The extraction of useful if-then rules from data based on statistical significance.

Decision trees are tree-shaped structures that represent sets of decisions. The decision tree approach can generate rules for the classification of a data set. Specific decision tree methods include Classification and Regression Trees (CART) and Chi Square Automatic Interaction Detection (CHAID). CART and CHAID are decision tree techniques used for classification of a data set. They provide a set of rules that can be applied to a new (unclassified) data set to predict which records will have a given outcome. CART typically requires less data preparation than CHAID (Lee, Siau, 2001).

CHAID modeling is an exploratory data analysis method used to study the relationships between a dependent measure and a large series of possible predictor variables those themselves may interact. The dependent measure may be a qualitative (nominal or ordinal) one or a quantitative indicator. For qualitative variables, a series of chi-square analyses are conducted between the dependent and predictor variables. For quantitative variables, analysis of variance methods are used where intervals (splits) are determined optimally for the independent variables so as to maximize the ability to explain a dependent measure in terms of variance components (Thearling, 2009).

4. The Study Design and Method

This study has three phases to measure and determine the operational risk factors that affected financial performance of firms. The study was conducted in Organized Industrial Region (OIB) of Ankara in 2008. The first phase consisted of the process of determining the levels of financial performance of firms and obtained data which used in third phase. Such data was obtained by means of financial analyses of balance sheets and income statements of companies available through Turkish Central Bank. Financial performance levels of firms were determined by ratio analysis. Table 1 shows the ratios and definitions that will be use in the first step of the study.

Table 1. Financial Variables

Ratios	Definition
Return on Equity	Net Income / Total Assets
Return on Assets	Net Income/ Total Equity
Profit Margin	Net Income/ Total Margin
Equity Turnover Rate	Net Revenues / Equity
Total Assets Turnover Rate	Net Revenues / Total Assets
Inventories Turnover Rate	Net Revenues / Average Inventories
Fixed Assets Turnover Rate	Net Revenues / Fixed Assets
Tangible Assets to Long Term Liabilities	Tangible Assets / Long Term Liabilities
Days in Accounts Receivables	Net Accounts Receivable/ (Net Revenues /365)
Current Assets Turnover Rate	Net Revenues/ Current Assets
Fixed Assets to Long Term Liabilities	Fixed Assets / Long Term Liabilities
Tangible Assets to Equities	Tangible Assets /Equities
Long Term Liabilities to Constant Capital	Long Term Liabilities / Constant Capital
Long Term Liabilities to Total Liabilities	Long Term Liabilities / Total Liabilities
Current Liabilities to Total Liabilities	Current Liabilities / Total Liabilities
Total Debt to Equities	Total Debt / Equities
Equities to Total Assets	Total Equity/Total Assets
Debt Ratio	Total Dept/Total Assets
Current Account Receivables to Total Assets	Current Account Receivables/ Total Assets
Inventories to Current Assets	Total Inventories / Current Assets
Absolute Liquidity	(Cash+Banks+ Marketable Sec.+ Acc. Rec.) / Current Liab.
Quick Ratio (Liquidity Ratio)	(Cash+Marketable Sec.+ Acc. Rec.)/ Current Liab.
Current Ratio	Current Assets/ Current Liabilities

Operational risk data which couldn't be access by balance sheets and income statements such financial management requirements, financial training and financial management skills of financial managers were collected in Phase 2. To collect data a questionnaire suitable for Turkish firms was designed and a field study was conducted in OIB. Field study covered

6.185 firms in OIB according to the records of OIB, but only 4.979 firms were found (others were closed or moved away). 2.630 firms (52,82 %) of 4.979 firms responded the questionnaire. Table 2 shows operational risk variables gained from firms by the way of survey.

Table 2. Operational Risk Variables

Operational Risk Variables
• Sector
• Legal Status
• Operating Length of Firms
• Number of Partners
• Number of Employees
• Annual Turnover
• Annual Balance Sheet
• Financing Model
• Usage of Credit
• Usage of Alternative Financing
• Technological Infrastructure
• Educational Background of Employees
• Educational Background of Managers
• Status of Managers
• Financial Education Background of Employees
• Financial Education Background of Managers
• Financial Training Need of Employees
• Financial Training Need of Managers
• Makers of Financial Strategies
• Expenditure of Energy
• Knowledge and Ability Levels of Workers on Financial Administration,
• Knowledge and Ability Levels of Workers on Financial Administration,
• Financial Problem Domains,
• Current Financial Risk Position.
• Knowledge About BASEL-II
• Knowledge About Quality Standards

In the third phase, qualitative and quantitative data obtained through phases 1 and 2 were analyzed by data mining. The main approach in analysis is discovering different operational risk levels and identifying the factors affected financial performance. Therefore, our study focused segmentation methods. By means of Chi-Square metrics CHAID is able to separately segment the groups classified in terms of level of relations. Therefore, leaves of the tree have not binary branches but as much branches as the number of different variables in the data. So, it was deemed convenient to use CHAID algorithm method in the study.

CHAID algorithms are developed on basis of two groups of variables, namely target variable and predictor variables that will explain the target variable. In the study financial performance is explained by means of operational risk variables of firms. Therefore, the financial performance indicator is considered as the target variable and operational risk variables (see Table 2) are considered as the predictor variables. CHAID algorithm is given below:

1. For each predictor variable X, find the pair of categories of X that is least significantly different (that is, has the largest p value) with respect to the target variable Y. The method used to calculate the p value depends on the measurement level of Y.
   1.1. If Y is continuous, use an F test.
   1.2. If Y is nominal, form a two-way cross tabulation with categories of X as rows and categories of Y as columns. Use the Pearson chi-squared test or the likelihood-ratio test.
   1.3. If Y is ordinal, fit a Y association. Use the likelihood-ratio test.
2. For the pair of categories of X with the largest p value, compare the p value to a prespecified alpha level, α_{merge}.
   2.1. If the p value is greater than α_{merge}, merge this pair into single compound category. As a result, a new set of categories of X is formed, and you start the process over at step 1.
   2.2. If the p value is less than α_{merge}, go on step 3.
3. Compute the adjusted p value for the set of categories of X and the categories of Y by using a proper Bonferroni adjustment.
4. Select the predictor variable X that has the smallest adjusted p value (the one that is most significant). Compare its p value to a prespecified alpha level, α_{split}
   4.1. If the p value is less than or equal to α_{split}, split the node based on the set of categories of X.
   4.2. If the p value is greater than α_{split}, do not split node. The node is a terminal node.
5. Continue the tree-growing process until the stopping rules are met (SPSS, 2001).

5. Results

In first phase of study, financial performance of firms were determined by ratio analysis and it was determined that 1836 (69.97 %) firms out of 2624 covered firms had good financial performance while 788 of them had poor financial performance. In other words 30.03 % of the covered firms financially distress.

SPSS Answer Tree software was used to develop the CHAID algorithm and $\alpha_{merge}=\alpha_{split}=$ 0.05 was taken as the basis to obtain the decision tree shown in Figure 1. 2.624 firms were categorized into 35 profiles on basis of operational risk factors by means of CHAID decision tree algorithm. These profiles give the operational risk level of the firms. As seen in Figure 1, according to these profiles it was determined that:

- educational background of managers (p<0.0001, Chi-square=64,60),
- number of employee (p<0.0000, Chi-square=46,32),

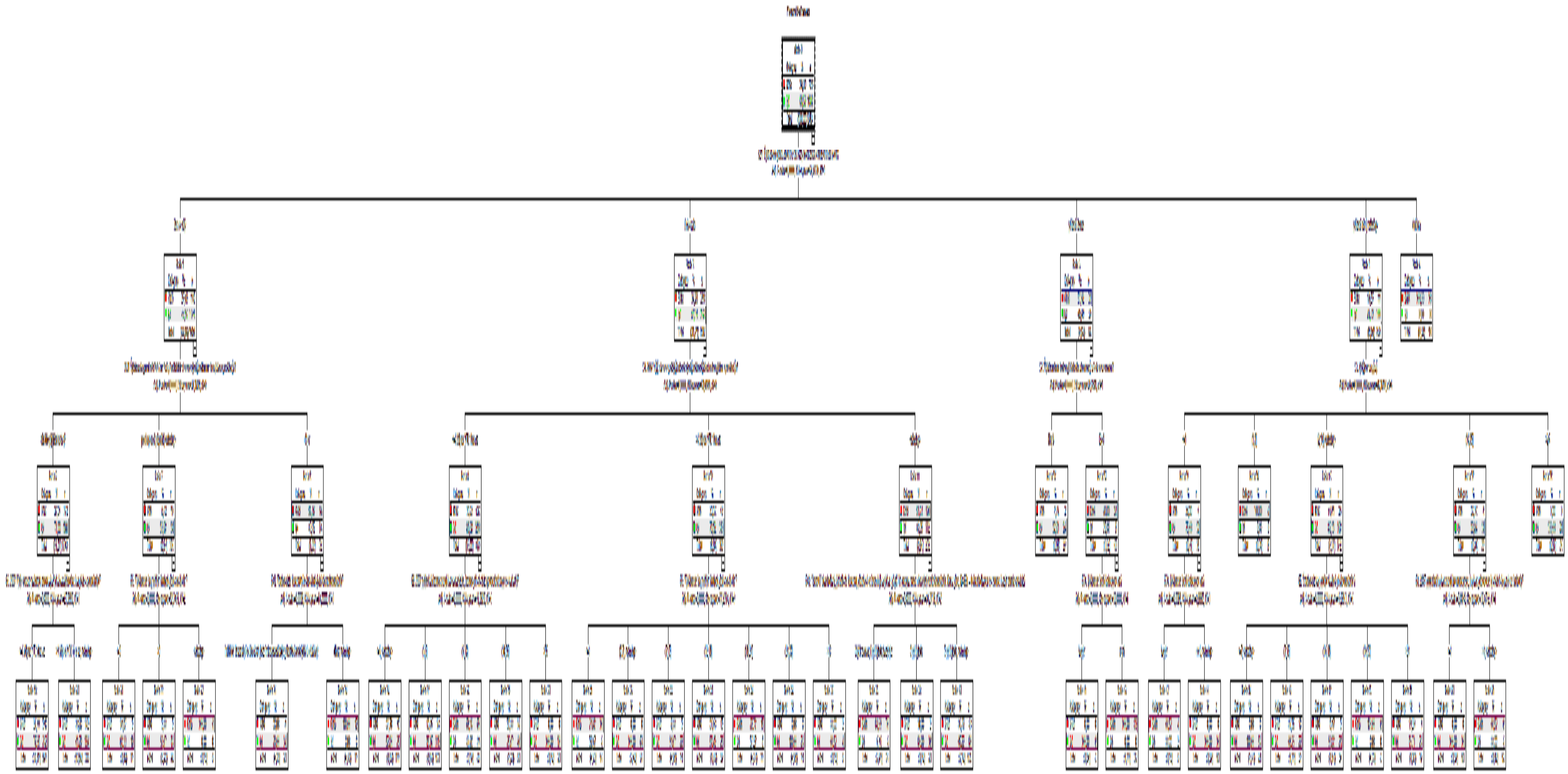

Figure1. Decision Tree obtained by CHAID Method

- status of managers ($p<0.000$, Chi-square=30,32),
- annual turnover ($p<0.0001$, Chi-square=27,06),
- operating length of firms ($p<0.0001$, Chi-square=62,34),
- makers of financial strategies ($p<0.0008$, Chi-square=14,93),
- expenditure of energy ($p<0.0001$, Chi-square=61,01),
- knowledge about Basel-II ($p<0.0001$, Chi-square=84,27),
- quality standards ($p<0.0001$, Chi-square=32,99),
- usage of credit ($p<0.0001$, Chi-square=40,00)

affected the financial performance of covered firms. In the other words, firms emphasize the management of these factors for hedging operational risk and raising financial performance. According to these profiles for reducing operational risk and improve financial performance, these factors should manage by the cover firms managers.

6. Conclusions

Risk management has become a vital topic for all institutions. It is thought that results of the study would be of great help for the firms to measure and hedging operational risk for reaching better financial performance. Some of expected contributions of using this methodology and CHAID Decision Trees Algorithm for hedging risk can be summarized as follows:

- to determine strategies for hedging risk,
- to prevent financial crises,
- to detect early warning sing for risk and crises,
- to identify road maps for better financial performance,
- to use resources more effectively,
- to ease burdens of banks by occurring strategically monitoring probability of risk factors being key characteristic for achieving BASEL II criteria.

According to this study results we suggested:

- educational background of managers should be at least vocational high school,
- number of employee should more then 25,
- managers of firms should be a professional,
- annual turnover of firms should be 4 million Turkish Liras,
- operating length of firms more then 20 years,

- financial strategies should make by board of management,
- expenditure of energy should be lower,
- knowledge about Basel-II should be high and firms should ready at least in 3 years,
- quality standards should be in firms,
- usage of credit should limited

to improve financial performance for the covered firms.

Acknowledgment

This paper is based on a project which was funded by The Scientific and Technological Research Council of Turkey (TUBITAK).

References

Basel Committee on Banking Supervision (2001) Working Paper on the Regulatory Treatment of Operational Risk, http://www.bis.org, accessed July, 2009.

Basel Committee on Banking Supervision (2003) The New Basel Capital Accord, Consultative Document, http://www.bis.org/publ/bcbsca.htm, accessed July, 2009.

Chang, S, Chang, H, Lin, C and Kao, S (2003) The Effect of Organizational Attributes on the Adoption of Data Mining Techniques in The Financial Service Industry: An Empirical Study in Taiwan, *International Journal of Management*, 20, 497-503.

Cornalbaa, C and Giudicib P (2004) Statistical Models for Operational Risk Management, *Physica A: Statistical Mechanics and its Applications,* 338(1-2),166-172.

Derby, B, L (2003) Data Mining for Improper Payments, *The Journal of Government Financial Management,* 52.

Eklund, T, Back, B, Vanharanta, H and Visa, A (2003) Using the Self- Organizing Map as a Visualization Tool in Financial Benchmarking, *Information Visualization,* 2,171-81.

Fayyad, U, Djorgovski, S.G and Weir, N (1996) Automating the Analysis and Cataloging of Sky Surveys in Fayyad, U, Piatetsky-Shapiro, G, Smyth, P, Uthurusamy, R (Eds), Advances in Knowledge Discovery and Data Mining*,* MIT Press, Cambridge, MA, 471-94.

Fayyad, U, Piatetsky-Shapiro, G and Smyth, P (1996) From Data Mining To Knowledge Discovery: An Overview", In Fayyad, U, Piatestsky-Shapiro, G, Smyth, P, Uthurusamy, R (Eds), Advances In Knowledge Discovery And Data Mining*, Mit Press*, Cambridge, MA.

Hoppszallern, S (2003) Healthcare Benchmarking, *Hospitals & Health Networks,* 77, 37-44.

Kleissner, C (1998) Data Mining for the Enterprise, Proceedings of the 31st Annual International Conference on System Sciences, IEEE Computer Society, 7, 295-304.

Kloptchenko, A., Eklund, T., Karlsson, J., Back, B., Vanhatanta, H., and Visa, A (2004) Combining Data and Text Mining Techniques for Analyzing Financial Reports, *Intelligent Systems in Accounting Finance and Management,* 12, 29-41.

Koyuncugil, A, S (2006) Fuzzy Data Mining and Its Application to Capital Markets, PHD. Thesis, Ankara University.

Koyuncugil, A, S and Ozgulbas, N (2006a) Is There a Specific Measure for Financial Performance of SMEs, *The Business Review, Cambridge,* 5(2), 314-319.

Koyuncugil, A, S and Ozgulbas, N (2006b) Financial Profiling of SMEs: An Application by Data Mining, *Proceedings of the European Applied Business Research (EABR) Conference*.

Koyuncugil, A, S and Ozgulbas, N (2006c) Determination of Factors Affected Financial Distress of SMEs listed in ISE by Data Mining, *Proceedings of 3rd Congress of SMEs and Productivity*, KOSGEB and Istanbul Kultur University, 159-170.

Koyuncugil, A S and Ozgulbas, N (2007) Detecting Financial Early Warning Signs in Istanbul Stock Exchange by Data Mining, *International Journal of Business Research*, VII (3).

Koyuncugil, A, S and Ozgulbas, N (2008a) Early Warning System Approach to SMEs Based on Data Mining as a Financial Risk Detector., 221-240, In Data Mining Applications for Empowering Knowledge Societies, Hakikur Rahman (Ed), Idea Group Inc., USA.

Koyuncugil, A, S, and Ozgulbas, N (2008b) Donor Research and Matching System Based on Data Mining in Organ Transplantation, *Journal of Medical Systems*, DOI 10.1007/s10916-008-9236-7.

Koyuncugil, A, S, and Ozgulbas, N (2009) Detecting Road Maps for Capacity Utilization Decisions by Clustering Analysis and CHAID Decision Trees, *Journal of Medical Systems*, DOI 10.1007/s10916-009-9258-9.

Kumar, N., Krovi, R. and Rajagopalan, B (1997) Financial Decision Support with Hybrid Genetic and Neural Based Modeling Tools, *European Journal of Operational Research,* 103, 339-349.

Lee, K, C, Han, I, and Kwon, Y (1996) Hybrid Neural Network Models for Bankruptcy Predictions, *Decision Support Systems,*18, 63-73.

Lee, S.J and Siau, K (2001) A Review of Data Mining Techniques, *Industrial Management & Data Systems,* 101(1), 41-46.

Magnusson, C, Arppe, A, Eklund, T and Back, B (2005) The Language of Quarterly Reports as an Indicator of Change in The Company's Financial Status, *Information & Management,* 42, 561-570.

Nazem, S and Shin, B (1999) Data Mining: New Arsenal for Strategic Decision Making, *Journal of Database Management*, 10, 39-42.

Piatetsky-Shapiro, G and Frawley, W.J (1991) Knowledge Discovery in Database, AAAI/MIT Press.

Ozgulbas, N and Koyuncugil, A, S (2006) Profiling and Determining the Strengths and Weaknesses of SMEs listed in ISE by the Data Mining Decision Trees Algorithm CHAID, *Proceedings of 10th National Finance Symposium*.

Ozgulbas, N and Koyuncugil, A, S (2009), Financial Profiling of Public Hospitals: An Application by Data Mining, *The International Journal of Health Planning and Management,* 24 (1), 69-83.

SPSS (2001) AnswerTree 3.0 User's Guide, USA: SPSS Inc.

Tam, K, Y and Kiang, M,Y (1992) Managerial Applications of Neural Networks: The Case of Bank Failure Predictions, *Decision Sciences,* 38, 926-948.

Thearling, K, Information About Data Mining and Analytic Technologies, http://www.thearling.com/text/dmwhite/dmwhite.htm, accessed July, 2009.

Weiss, S.H and Indurkhya, N (1998) Predictive Data Mining: A Practical Guide, Morgan Kaufmann Publishers, San Francisco, CA.