Unearthing Top 3 Business Strategies using Data Mining Techniques

Khalid Mehmood Iraqi, PhD
Dean of Management and Administrative Sciences, University of Karachi, 75270, Karachi, Pakistan

Huda Yasin
Karachi University Business School (KUBS), University of Karachi, 75270, Karachi, Pakistan

Mohsin Mohammad Yasin
Institute of Business Management (IoBM), Karachi, Pakistan

ABSTRACT
Business strategies portray the measures that should be taken with the intention of achieving enduring objectives. Above and beyond, enduring objectives characterize the result anticipated from taking up particular strategies. Nevertheless, opting strategies which go well with organization is not an undemanding task. In this research paper, on the basis of diverse organizations' data, a novel methodology to get top 3 strategies for a business is presented. For this purpose, a dummy dataset of different organizations has been generated. The dummy dataset contains 134 influential variables as well as the successful strategies adopted by the considered organizations. Two different similarity measures namely, Jaccard coefficient and Dice coefficient have been applied. Besides, Pearson correlation coefficient is also applied on the dummy dataset. It is predicted that by means of our novel approach, a business strategist would obtain the suitable business strategies for his or her organization in an efficient and quite tranquil way.

Keywords
Business analytics, Data mining, Decision support system

1. INTRODUCTION
Business Intelligence (BI) has drastically benefitted with Data Mining which entails assessing big data to come across pertinent patterns for better understanding and decision making in a specific business framework. Besides, big data permits organizations to dig out value from great volumes of data [1] [2]. And, organizations that successfully design and implement big data strategies stand to get a competitive advantage [3]. There are several different forms of analysis that can be carried out so as to retrieve information from big data [4]; in addition to this, diverse effects and outcome are obtained from different types of analysis. Furthermore, the type of data mining technique one should apply actually depends on the type of business problem that one is attempting to disentangle.

According to David (2010), it is not solely the duty of top executives to make a strategy [5]. In the strategic planning process, both middle and lower level managers are engaged [6]. It is vital to understand that to help ensure synchronization, facilitation, and obligation while avoiding discrepancies as well as inefficiencies, all individuals at the different stages which are accountable for strategic planning superlatively partake as well as grasp the strategies at the other levels of the organization. To achieve enduring objectives, business strategy enlightens the actions that should be taken [7]. And enduring objectives portray the outcome projected from taking up a particular strategy. A number of large diversified organizations use a blend of two or more strategies in a simultaneous manner; however a blend or grouping strategy can be very uncertain if taken very far. Moreover, not a single organization can meet the expense of carrying all the strategies that might benefit the business or an organization.

Different research studies have been carried out to analyze different business patterns by means of data mining techniques. Hancu (2008) proposed a technique to support merging decisions [8]. His technique was derived from the financial statements analysis as well as the utilization of web logs taken out from the application of multi-server search. The aim of his study was to provide assistance to the companies while making a decision to merge with any other company, or to make them alert about the reality that a rival or opponent could combine in the upcoming period. According to his proposed model, if an entity shares a similar activity code with the initial entity, or possesses an associated code, the model would suggest to merge with that entity. On the other hand, it has been recognized that mergers and acquisitions do not occur in an unvaryingly distributed manner, but have a tendency to mount up across a range of dimensions - a fact often named as merger waves. Sziics (2013) analyzed the clustering properties of merger waves [9]. On the basis of approximately 600,000 individual transactions, he studied the level of agglomeration of acquisition activity surrounded by clusters of geographic, temporal and industrial nearness. According to his conclusion, important clustering comes about in time and across industries, whilst the consequences on geographic clustering are assorted. This encourages the notion that by and large merger waves are driven via neoclassical motives.

However, a good piece of research has been done in this domain, but in this research paper, a novel approach is proposed to get top 3 strategies for a business. For this purpose, a dummy dataset of different organizations has been generated. The dummy dataset contains 134 influential variables as well as the successful strategies adopted by the considered organizations. On the basis of adopted business strategy, the dummy dataset is categorized into four different clusters. Initially Jaccard coefficient, a similarity measure, would assess the similarities between the input data and the previously stored dataset. In the beginning, the considered variables only belong to internal and external strategic position of the organizations. After that, it would link the input data to a particular cluster having the highest similarity score. Subsequently, top 3 business strategies will be obtained on the basis of Dice coefficient, another similarity measure, and Pearson correlation coefficient. The adopted methodology, results and thorough discussion is presented in the following sections.

2. METHODOLOGY

This section will provide information concerning the methods used in this research. Also, the structural plan which will be followed to achieve the main objective is explained. With the intention of achieving the research objectives, the organizations' dataset is categorized into four different clusters named as cluster I – conservative, cluster II – aggressive, cluster III – defensive, and cluster IV – competitive. This categorization is based on the type of the successful strategy adopted by the organizations namely, conservative, aggressive, defensive, and competitive. Now, the dummy dataset can be scrutinized in connection with these four clusters. By the strategist, the input data will be obtained and then it will be compared with the organizations' data resides in these four clusters. But initially in this comparison, only those factors are considered which are included in internal and external strategic position, and afterward the comparison is made by means of similarity measure. After assessing the similarities, the input data will move into any of the predefined clusters having the highest similarity score. Besides, a criterion will also be discussed if cluster II, III, or all the four clusters show the highest similarities with the input data. In order to determine precise and specific business strategies, further data mining techniques including similarity measure and Pearson correlation coefficient will be applied on rest of the 102 variables.

2.1 Research Approach and Design

According to Balnaves and Caputi (2001), quantitative approach and qualitative approach are the two chief types of research approaches [10]. In this research paper, a quantitative research has been conducted as the presented methodology is based on experiments.

In this research paper, the novel approach is that data mining techniques, a mushrooming field, have been applied to the dummy dataset. The applied data mining techniques are: similarity measure and Pearson correlation coefficient. After applying data mining techniques on the dummy dataset, interesting patterns and domino effects have been analyzed. Besides, the projected structural plan of this research study is presented in Figure 1.

2.2 An Overview of Chosen Data Mining Techniques

In accordance with several researchers, there are numerous diverse techniques used to perform data mining tasks [11] [12] [13] [14]. Furthermore, such methods not merely require particular data structures types, but entail particular types of algorithmic methods too [15]. For this research work, two different data mining techniques namely, similarity measure and Pearson correlation coefficient have been chosen. These techniques are elucidated below;

2.2.1 Similarity Measure

In the context of internet searches using a search engine, the use of similarity measures is not new [15]. However, it has been analyzed that in the domain of business data, this notion is not too old.

According to Dunham (2003), Dice, Jaccard, Cosine, and Overlap are some of the common similarity measures [15]. And in this research work, the similarity measures which have been employed are Jaccard coefficient and Dice coefficient;

$$Jaccard\ Coefficient = \text{sim}(B_y, B_z) =$$

$$\frac{\sum_{j=1}^{i} B_{yj}B_{zj}}{\sum_{j=1}^{i} B_{yj}^2 + \sum_{j=1}^{i} B_{zj}^2 - \sum_{j=1}^{i} B_{YJ}B_{ZJ}} \quad (1)$$

$$Dice\ Coefficient = \text{sim}(B_y, B_z) = \frac{2\sum_{j=1}^{i} B_{yj}B_{zj}}{\sum_{j=1}^{i} B_{yj}^2 + \sum_{j=1}^{i} B_{zj}^2} \quad (2)$$

2.2.2 An Overview of Pearson's Correlation Analysis

According to Weiss (1998), the Pearson correlation coefficient r can be defined in this way [16]: consider H and K are the two variables; H_1, H_2,…, H_y are the y values contained by the variable H, whereas K_1, K_2,…, K_y are the y values contained by the variable K. Now, suppose $\bar{H}$ and $\bar{K}$ are the mean of H and K correspondingly. Pearson correlation coefficient r is;

$$r = \frac{\sum(H_i - H)(K_i - \bar{K})}{\sqrt{\sum(H_i - \bar{H})^2 \sum(K_i - \bar{K})^2}} \quad (3)$$

3. DATA DESCRIPTION

The dummy dataset is comprised of total 134 variables. The dataset contains numeric, real and categorical data fields. A thorough view of the chosen variables is presented in table 1-6.

In addition, table 1 contains variables which represent the internal strategic position (financial position and competitive position) as well as external strategic position (stability position and industry position) of an organization. These variables are 32 in total. Variables which are dealing with integration strategies (including forward, backward, as well as horizontal integration strategy) are presented in table 2. There are 9 variables which are involved in forward integration, 10 variables in backward integration, and 7 variables in horizontal integration. Product development, market penetration, and market development are the chief types of intensive strategies; the related variables are presented in table 3. And the table clearly shows that market penetration contains 6 variables, market development contains 10 variables, and 5 variables are included under product development. Table 4 shows variables associated with diversification strategies; both related diversification and unrelated diversification variables are presented in this table. Related diversification contains 6 variables and unrelated diversification contains 11 variables. Retrenchment, divestiture, and liquidation are the main categories of defensive strategies as shown in table 5; the variables associated with retrenchment, divestiture, and liquidation are 10, 6, and 3 respectively. Table 6 gives a clear idea of variables which present resource based view of an organization. In this table, the three main resources, namely; tangible, human, and organizational resources contain 5, 7, and 7 organizations' specific resources respectively.
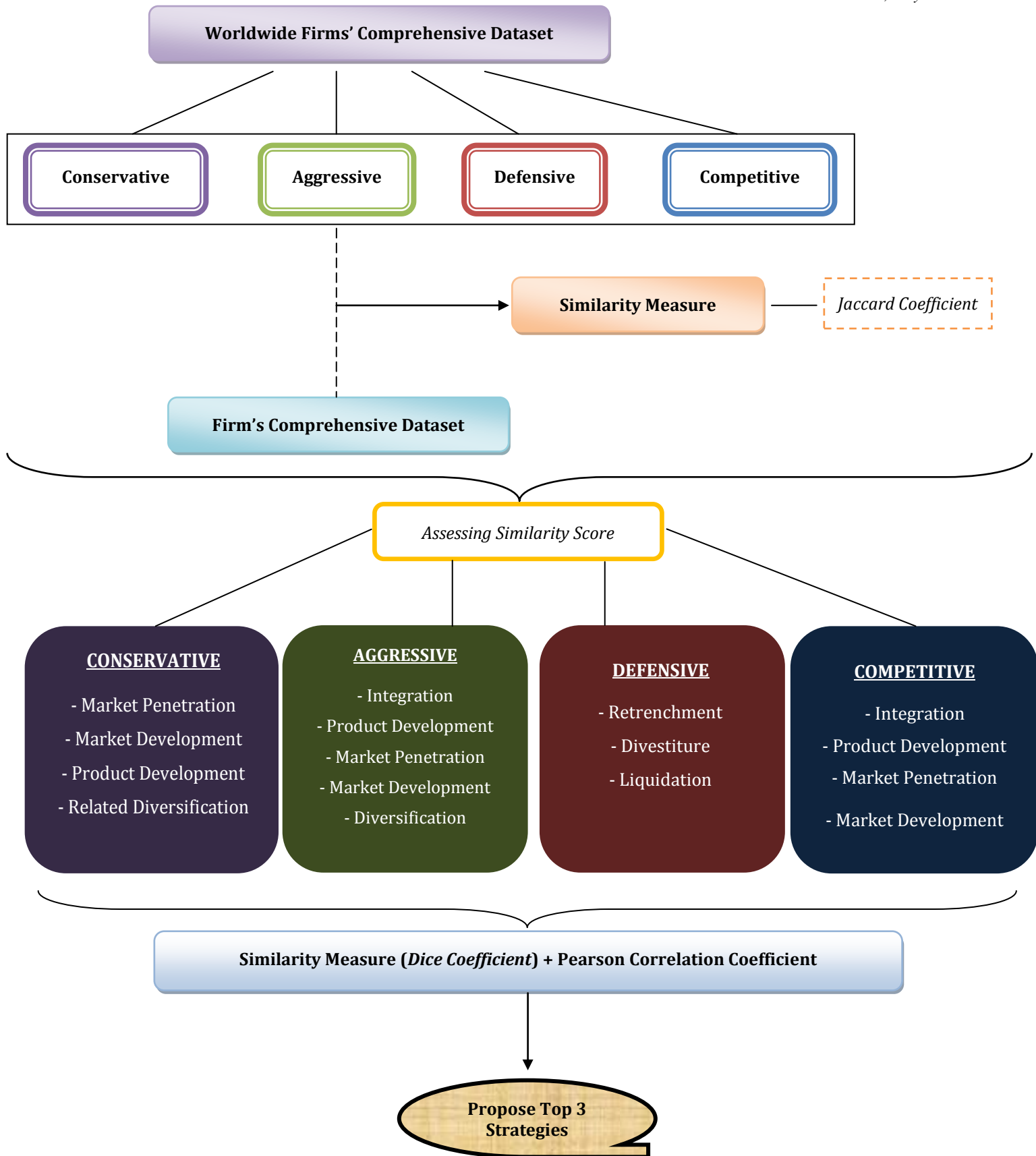
Worldwide Firms' Comprehensive Dataset

| Conservative | Aggressive | Defensive | Competitive |

Similarity Measure — *Jaccard Coefficient*

Firm's Comprehensive Dataset

Assessing Similarity Score

CONSERVATIVE

- Market Penetration
- Market Development
- Product Development
- Related Diversification

AGGRESSIVE

- Integration
- Product Development
- Market Penetration
- Market Development
- Diversification

DEFENSIVE

- Retrenchment
- Divestiture
- Liquidation

COMPETITIVE

- Integration
- Product Development
- Market Penetration
- Market Development

Similarity Measure (*Dice Coefficient*) + Pearson Correlation Coefficient

Propose Top 3 Strategies

Figure 1. Structural Design

4. RESULTS AND DISCUSSION

Firstly, the input data is compared with the already stored dataset (dummy dataset). This comparison is made on the basis of Jaccard coefficient, a similarity measure. At the moment, the similarity measure is applied on the internal strategic factors and external strategic factors. Our considered internal and external strategic factors are 32 in total and are presented in table 1. After observing the entire outcome, it can be stated that the input data gets the highest similarity score with cluster 2 – Aggressive. Similarity score is presented in table 7. So, the input dataset moved into cluster 2. In addition, if the input data also gets the highest similarity score with cluster 1 or cluster 3 or cluster 4, or if the input data gets the highest similarity score with all the four clusters then the same procedure will be applied to them as well. Now, the main task is to find out a particular business strategy.

Chiefly, aggressive strategy comprises of 5 different strategies namely, Integration, Product Development, Diversification, Market Development and Market Penetration. Hence, now the task is to get top 3 strategies from the mentioned 5 strategies. Nevertheless, some of these strategies contain different sub strategies as well. Forward, backward, and horizontal integration are linked with integration strategy. Product development, market penetration, and market development are the types of intensive strategy. Diversification strategy comprises of related diversification strategy and unrelated diversification strategy. So in point of fact, the task is to get top 3 strategies from the mentioned 8 strategies.

To get an accurate result, those variables which are involved while analyzing resource based view of organizations are also considered. It has been scrutinized that either the tangible resources (plant along with equipment, locality, unrefined materials, technology, and machineries), human resources (workers, training, know-how, brainpower, knowledge, expertise, and aptitude), and organizational resources (structure of the firm, planning procedures, information systems, official documents, trade names, exclusive rights, and databases) of the input data are valuable, rare, pricey to emulate, and organized so as to get a sustained competitive advantage. In order to achieve our main objective, Pearson correlation coefficient has been applied on the variables having interval and ratio scale; and Dice coefficient on rest of the variables. Next, two numbers are added, i.e. Pearson correlation coefficient, and similarity measure score (Dice coefficient). The top 3 business strategies are chosen on the result of this added score. As a result, the proposed top 3 strategies for the input data are related diversification (having 91.3% final score), market penetration (having 86.7% final score), and market development (having 83.2% final score). Table 8 contains the final values, plus they are presented in percentages. As a final point, for the input data, it has been scrutinized that the business strategies proposed by our approach and the strategies planned by a strategist are similar.

5. CONCLUSION AND FUTURE WORK

Choosing effective business strategies is not only a tedious task but also takes so much time of a business strategist. In this research paper, a novel methodology is presented in order to get top 3 strategies for a business on the basis of diverse organizations' data. For this purpose, a dummy dataset of different organizations has been generated. Similarities between the input data (testing data) and the previously stored dataset (training data) are evaluated. Initially, the considered variables only belong to internal and external strategic position of the organizations. Based on the highest similarity score, the input data will be linked to a particular cluster. Afterward, top 3 business strategies will be obtained on the basis of correlation and similarity measure (excluding the variables linked with internal and external strategic position). It has been analyzed that by means of the presented novel approach, a business strategist would obtain the suitable business strategies for his or her organization in an efficient manner.

Many different aspects would be analyzed in future study. First of all, a real dataset of diverse organizations would be gathered. Secondly, this whole methodology would be automated. After that, K-nearest neighbor strategy and different hierarchical clustering algorithms would be applied to analyze other interesting patterns.

In addition, some more influencing variables would also be included to get more efficient and optimized result.

6. ACKNOWLEDGEMENT

Table 1: Factors Involved in Internal and External Strategic Position

Strategic Position – Internally	
Financial Circumstances	*Competitive Circumstances*
ROI: Return on Investment	Market Share
EPS: Earnings Per Share	Exploitation of Capacity
Etc.	*Etc.*
Strategic Position - Externally	
Stability Circumstances	*Industry Circumstances*
Changes in Technology	Potential of Expansion
Inflation Rate	Potential of Getting Profit
Etc.	*Etc.*

Table 2: Integration Strategies

Forward Integration
Existing Distributors Costly
Existing Distributors Reliability
Etc.

Backward Integration
Existing Suppliers Expensive
Existing Suppliers Reliability
Etc.

Horizontal Integration
Organization Can Gain Monopolistic Distinctiveness
Challenge via Centralized Government
Etc.

Table 3: Intensive Strategies

Market Penetration
Existing Market Saturated
Customers Usage Rate Boost Up
Etc.

Market Development
Novel Channels of Distribution Accessible
Novel Channels of Distribution Trustworthy
Etc.

Product Development
Organization Having Thriving Developed Products
Organization Competes in Fast Technological Developments
Etc.

Table 4: Diversification Strategies

Related Diversification
Organization Competes in Lose-pace Growth Industry
New Related Products Would Drastically Enhance Existing Products Sales
Etc.

Unrelated Diversification
Latest Unrelated Products Would Enhance Revenues
Organization Competes in an Extremely Competitive Industry
Etc.

Table 5: Defensive Strategies

Retrenchment
Has a Clear Unique Aptitude But Failed Constantly
Organization is a Weaker Competitor
Etc.

Divestiture
Failed Retrenchment Strategy
Division Requires More Resources
Etc.

Liquidation
Failed Retrenchment and Divestiture Strategy
Bankruptcy Only Substitute
Etc.

Table 6: Resource Based View

Organization's Main Resources	Organization's Specific Resources	Valuable	Rare	Pricey to Emulate	Organized	Sustained Competitive Advantage
Tangible Resources	Plant along with Equipment Locality Unrefined Materials Technology Machineries			*Yes/ No*		
Human Resources	Workers Training Know-how Brainpower Knowledge Expertise Aptitude			*Yes/ No*		
Organizational Resources	Structure of the Firm Planning Procedures Information Systems Official Documents Trade Names Exclusive Rights Databases			*Yes/ No*		

Table 7: Jaccard Coefficient Similarity Score

	Cluster I - Conservative	Cluster II - Aggressive	Cluster III - Defensive	Cluster IV - Competitive
Firm's Input Data	77.36%	88.28%	81.73%	63.42%

Table 8: Top 3 Business Strategies

	Forward Integration	Backward Integration	Horizontal Integration	Market Penetration	Market Development	Product Development	Related Diversification	Unrelated Diversification
Firm's Input Data	58.6%	64.2%	55.4%	**86.7%**	**83.2%**	77.3%	**91.3%**	72.1%

7. REFERENCES

[1] Mayer-Schönberger, V. & Cukier, K. (2014). Big Data: A Revolution That Will Transform How We Live, Work, and Think, Eamon Dolan/Mariner Books

[2] Davenport, T.H. (2014). Big Data at Work: Dispelling the Myths, Uncovering the Opportunities, *Harvard Business Review Press*.

[3] Parise, S. Iyer, B., & Vesset, D. (2012). Four Strategies To Capture And Create Value From Big Data. Ivey Business Journal.

[4] Kantardzic, M. (2003). *Data Mining: Concepts, Models, Methods, and Algorithms*. John Wiley & Sons.

[5] David, F. R. (2010). *Strategic Management*. Prentice Hall, 13th Edition.

[6] Boone, L.E., & Kurtz, D.L. (2011). *Contemporary Business*, Wiley; 14th edition.

[7] Robbins, S.P. & Coulter, M. (2010). *Management*. Prentice Hall; 11th edition.

[8] Hâncu, L. (2008). Data-Mining Techniques for Supporting Merging Decisions. Int. J. of Computers, Communications & Control, *Suppl. issue: Proceedings of ICCCC 2008*, vol. III, 322-326.

[9] Szucs, F. (2013). *Clustering properties of merger waves: space, time or industry?* DIW Berlin

[10] Balnaves, M., & Caputi. (2001). *Introduction to Quantitative Research Methods: An Investigative Approach*. SAGE Publications.

[11] Shmueli, G., Patel, N. R., & Bruce, P.C. (2010). *Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner*, Wiley; 2nd edition.

[12] Linoff, G.S. & Berry, M. J. (2011). *Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management*. Wiley; 3rd edition.

[13] Witten, I. H., Frank, E., & Hall, M.A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann; 3rd edition.

[14] Yin, Y., Kaku, I., Tang, J., & Zhu, J. (2011). *Data Mining: Concepts, Methods and Applications in Management and Engineering Design (Decision Engineering)*. Springer.

[15] Dunham, M. H. (2003), *Introduction to Data Mining (Introductory and Advanced Topics)*. Pearson Education India.

[16] Weiss, N.A. (1998). *Elementary Statistics*, pp. 733-777