APPLICATION OF ASSOCIATION RULES IN DATA MINING FOR FINANCIAL DECISION MAKING

Vo Ngoc Tuyet Phuong
Systemgear Vietnam ltd, Vietnam
vntphuong.khtn@gmail.com
Huynh Van Tran
Larion Computing ltd, Vietnam
huynhvantran@gmail.com
Do Phuc
University of Information Technology - HCM, Vietnam
phucdo18@gmail.com

Abstract

This objective of this study is to discover the association rules that express the relationship of financial ratios. The relationship of financial ratios is very useful because if we know a ratio, we can predict other ratios through its relationship. We collected data set from financial statements of real estate enterprises on the stock exchanges in Vietnam in the period 2011 to 2013. Preprocessing has been done to convert the data set from financial statements to financial ratios. We used association rule mining method to extract association rules of financial ratios. These rules showed that there is a significant positive relationship between accounts receivable turnover ratio, inventory turnover ratio, and total asset turnover; there is a negative relationship between these ratios and profit margin ratio. The results proved that we can predict this year's profit margin ratio base before year's profit margin ratio, inventory turnover ratio, accounts receivable turnover ratio. This relationship is useful for internal and external financial data users in making their economic decisions, including investing and performance evaluation decisions.

Keywords*: data mining, association rule, financial ratios, real estate, Weka*

INTRODUCTION

Nowadays, there is large amount of data on the internet. Analyzing this large amount of data and extracting useful information from it is necessary. Data Mining is to extract the knowledge from the large data set. This information can be used for any of the following applications: market analysis, fraud detection, customer retention, production control, science exploration…

Financial ratios are one of the most common tools of managerial decision making. Financial ratios analysis is a key business skill for any business owner or investors. Financial ratios illustrate the strengths and weaknesses of a business. These ratios help them compare the performance of companies that are very different in size.

Since the rapidly growing volume of financial ratios has far exceeded our ability to analyze them manually. There is a critical need for automated approaches to effective and efficient utilization of massive financial data to support companies and individuals in strategic planning and investment decision-making(Zhang, 2004).

In this paper, we collected data fromfinancial statements of real estate enterprises on the stock exchanges in Vietnam in the period 2011 to 2013. Preprocessing has been done to convert the data on financial statements to financial ratios. We used association rule mining method in data mining to extract association rules of financial ratios. The relationship of

ISBN 978-602-19725-8-8

financial ratios is very useful because if we know ratio or a few ratios, we can predict other ratios through its relationship in association rules.

This paper is organized as follows: A brief introduction is presented in section 1. In section 2, we introduce about association rule mining method in data mining, Weka software, the choice of financial ratios. The model is developed and tested in section 3. Section 4 reviews empirical results obtained from this model. Finally, in this section 5, we highlight a number of challenges and trends for future research in this area.

THEORETICAL FOUNDATIONS

ASSOCIATION RULES

Basics objective of finding association rules is to find allco-occurrence relationship called associations. Since it was first introduced in 1993 by Agrawal et al, it has attracted a greatdeal of attention.Many efficient algorithms, extensions and applications have beenreported. The classic application of association rule mining is market basket data analysis, which aims to discover how items purchased by customers in a supermarket (or store) are associated. Association rules are of form X→ Y, where X and Y are collection of items and intersection of X and Y is null [15]

The problem of mining association rules can be stated as follows: Let I = { i_1, i_2...i_m } be a set of items. Let T = (t_1, t_2...t_n) be a set of transactions (the database), where eachtransaction$t_i$is a set of items such that$t_i \subseteq$ I. Anassociation ruleis an implication of theform:

$$\boldsymbol{X \rightarrow Y, where\ X \subset I, Y \subset I, and\ X \cap Y = \emptyset}$$

XorY is a set of items, called item sets.

A transaction $t_i \in$Tis said tocontainan itemsetXifXis a subset oft_i(we also say thatthe itemsetXcoverst_i). Thesupport countofXinT (denoted by X.Count) is the number of transactions in T that contain X. The strength of a rule is measured by its support and confidence. [15]

Support: The support of a rule X→Y is thepercentage of transactions inTthat containsX∪Y, and can be seen as an estimate of the probability, Pr (X∪Y). The rule support thusdetermines how frequent the rule is applicable in the transaction setT. Letnbe the numberof transactions inT. The support of the ruleX→Yis computed as follows:

$$\mathrm{Support} = \frac{(\mathrm{X} \cup \mathrm{Y}).\mathrm{Count}}{\mathrm{n}}$$

Support is a useful measure because if it is too low, the rule may just occur due to chance.Furthermore, in a business environment, a rulecoveringtoo few cases (or transactions)may not be useful because it does not make business sense to act on such a rule (notprofitable). [15]

Confidence: The confidence of a rule,X→Y, is the percentage of transactions in T that containXalso containY. It can be seen as an estimate of the conditional probability, Pr (Y|X). It is computed as follows:

$$\mathrm{Confidence} = \frac{(\mathrm{X} \cup \mathrm{Y}).\mathrm{Count}}{\mathrm{X}.\mathrm{Count}}$$

Confidence thus determines thepredictabilityof the rule. If the confidence of a rule is too low, one cannot reliably infer or predictYfromX. Arule with low predictability is of limited use. [15]

ISBN 978-602-19725-8-8

For example, the information that customers who purchase computers also tend to buy antivirus software at the same time is represented in Association Rule:

Computer =>antivirus software [support = 2%; confidence = 60%].

Rule support and confidence are two measures of rule interestingness. They respectively reflect the usefulness and certainty of discovered rules. A support of 2% for Association Rule means that 2% of all the transactions under analysis show that computer and antivirus software are purchased together. A confidence of 60% means that 60% of the customers who purchased a computer also bought the software. Typically, association rules are considered interesting if they satisfy both a minimum support threshold and a minimum confidence threshold. Such thresholds can be set by users or experts in a particular domain of application. Additional analysis can be performed to uncover interesting statistical correlations between associated items, Han, J. (2001)

FINANCIAL RATIOS

Financial ratios are useful as indicators of a firm's performance and financial situation. Most ratios can be calculated from information provided by the financial statements. Financial ratios can be used to analyze trends and to compare the firm's financials to those of other firms. In some cases, ratio analysis can predict future bankruptcy [16]

We use 4 types of ratios: liquidity ratios, asset turnover ratios, financial leverage ratios, profitability ratios. We choose following ratios to research.

TABLE 1: TABLE OF FINANCIAL RATIOS

No	Ratio	Definition
1	Net Income/sales	(Net Profit)/main stream
2	Net Income/assets	(Net profit)/total assets amount
3	Earnings before Interest and taxes/assets	(main stream profit + other stream profit+invest profit)/total assets amount
4	Net income before extraordinary items/assets	(main stream profit + other stream profit)/ net stock amount
5	Net income before extraordinary items/stockholders' equity	(main stream profit + other stream profit)/ net stock amount
6	Cash/current liabilities	Total cash flow / short term loan
7	Sales/ Assets	Main stream income/total assets amount
8	Cost of goods sold/inventory	Cost of sale item/(work in progress +stock amount)
9	Account receivable/sales	Receivable inv/main stream income
10	Liabilities/stockholder's equity	(float debt total + long term loan)/net stock amount
11	Assets/stockholder's equity	Total assets amount /net stock amount
12	Long term debt/assets	Long term loan/ total assets amount
13	Liabilities / Assets	(float debt total + long term loan)/total assets amount
14	Current assets/current liabilities	Short term deposit / short term loan

ISBN 978-602-19725-8-8

15	Quick assets/ current liabilities	(total asset amount – long term asset)/ short term loan
16	Cash / assets	(total cash flow/ total assets amount)

WEKA SOFTWARE

Weka is a collection of machine learning algorithms for data mining tasks. Weka has been developed by The University of Waikato, New Zealand. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes [14]

FIGURE 1: WEKA SOFTWARE

Basic Functionality (Figure 2)

Here is a summary of WEKA's main feature, Bouckaert (2010):

Data preprocessing: As well as a native file format (ARFF), WEKA supports various other formats (for instance CSV, Matlab ASCII files), and database connectivity through JDBC.Data can be filtered by a large number of methods (over 75), ranging from removingparticular attributes to advanced operations such as principal component analysis.

Classification: One of WEKA's drawing cards is the more than 100 classification methods it contains. Classifiers are divided into "Bayesian" methods (Naive Bayes, Bayesian nets, etc.), lazy methods (nearest neighbor and variants), rule-based methods (decision tables, OneR, RIPPER), tree learners (C4.5, Naive Bayes trees, M5), function-based learners (linear regression, SVMs, Gaussian processes), and miscellaneous methods. Furthermore, WEKA includesmeta-classifiers like bagging, boosting, stacking; multiple instance classifiers; and interfaces for classifiers implemented in Groovy and Jython

Clustering: Unsupervised learning is supported by several clustering schemes, including EM-based mixture models-means, and various hierarchical clustering algorithms. Though not as many methods are available as for classification, most of the classic algorithms are included

Association rules: Association Rule mining is important task of data mining that finds the probability of co-occurrence of items in a collection. The major goal is to extract interdependence associations' structures among the item sets in the transaction databases or other data repositories.

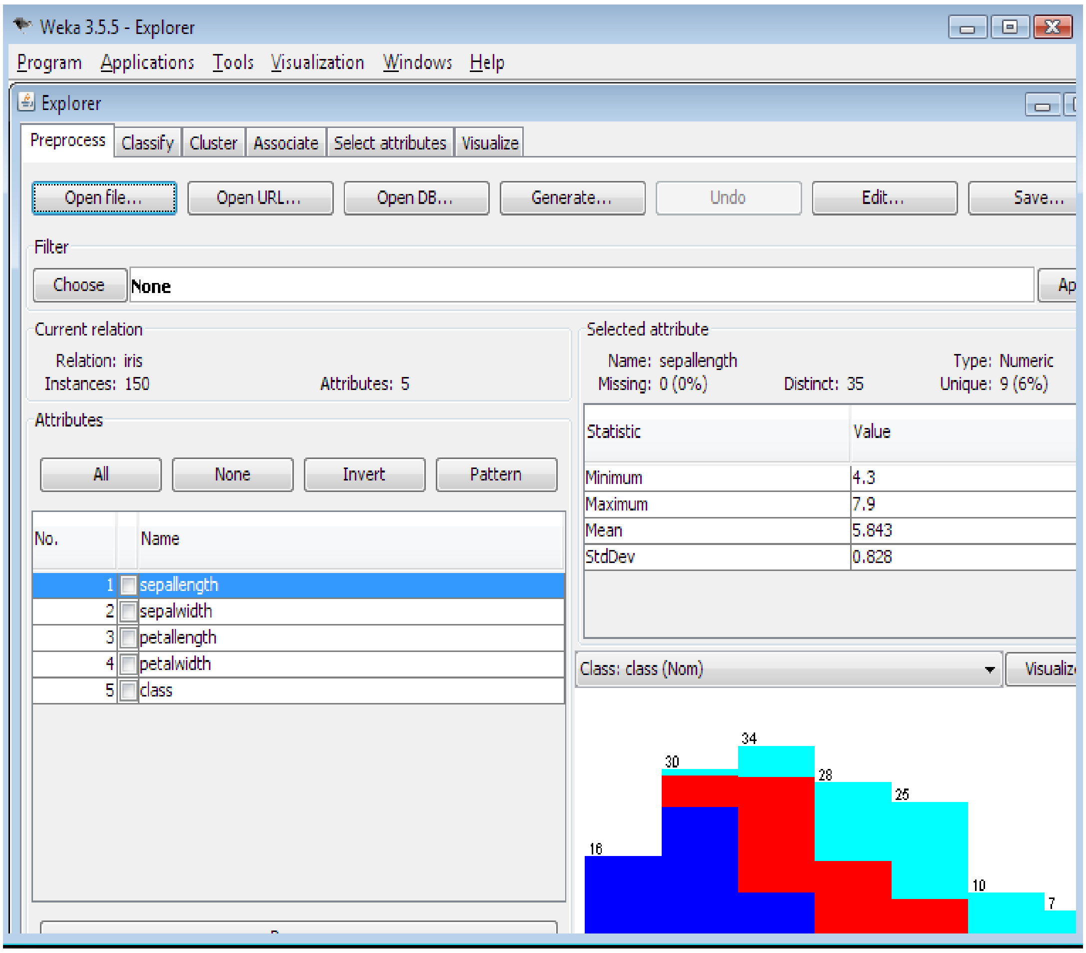
Attribute selection: The set of attributes used is essential for classification performance. Various selection criteria and search methods are available.

Data visualization: Data can be inspected visually by plotting attribute values against the class, or against other attribute values. Classifier output can be compared to training data in order to detect outliers and observe classifier characteristics and decision boundaries. For specific methods there are specialized tools for visualization, such as a tree viewer for any

ISBN 978-602-19725-8-8

method that produces classification trees, a Bayes network viewer with automatic layout, and a dendrogram viewer for hierarchical clustering.

FIGURE 2: WEKA EXPLORER



Research Methodology

Collecting Data Set

We collected financial statements (balance sheet, income statement) of 58 real estate enterprises on the stock exchanges in Vietnam in the period 2011 to 2013 (at web site http://cophieu68.vn)

ISBN 978-602-19725-8-8

FIGURE 3: DATA SET OF FINANCIAL STATEMENTS

	A	B	C	D	E	F
1	CAN DOI KE TOAN	2013	Q4 2013	Q2 2013	Q1 2013	2012
2	Tài sản ngắn hạn###Current Assets	120389	120132	136824	129603	105074
3	Tiền và các khoản tương đương tiền###Cash and Cash Euivalents	1467	1467	3458	2600	3344
4	Các khoản đầu tư tài chính ngắn hạn###Short term financial investment	0	0	0	0	0
5	Các khoản phải thu ngắn hạn###Short term Account Receivables	42021	44050	50613	61717	51875
6	Hàng tồn kho###Inventory	63548	63548	72262	57991	46267
7	Tài sản ngắn hạn khác###Other Current Assets	13353	11067	10492	7296	3588
8	Tài sản dài hạn###Non-current Assets	256935	260650	256026	260016	269248
9	Các khoản phải thu dài hạn###Long term Account Receivable	0	0	0	0	0
10	Tài sản cố định ###Fixed assets	248060	249901	249971	252239	261168
11	Tài sản cố định hữu hình - Giá trị hao mòn lũy kế	-200536	-201358	-192402	-190883	-186409
12	Tài sản cố định thuê tài chính - Giá trị hao mòn lũy kế	0	0	0	0	0
13	Tài sản cố định vô hình - Giá trị hao mòn lũy kế	-98	-98	-87	-81	-75
14	Lợi thế thương mại	0	0	0	0	0
15	Bất động sản đầu tư ###Real Estate Investment	0	0	0	0	0
16	Các khoản đầu tư tài chính dài hạn###Long term Finacial Investments	880	2562	2562	4262	4262
17	Tài sản dài hạn khác###Other long term assets	7995	8186	3492	3515	3818
18	TỔNG CỘNG TÀI SẢN ###TOTAL ASSETS	377324	380782	392850	389620	374323
19	Nợ phải trả ###Liabilities	364051	367280	374695	364280	341115
20	Nợ ngắn hạn###Short term Liabilities	238651	241880	221852	208275	193501
21	Nợ dài hạn###Long term Liabilities	125400	125400	152843	156005	147614
22	Dự phòng nghiệp vụ	0	0	0	0	0
23	No khac	0	0	0	0	0
24	Vốn chủ sở hữu ###Owners equity	13273	13502	18155	25340	33208
25	Nguồn kinh phí và quỹ khác###Expenditures and Other Funds	0	0	0	0	0
26	Vốn chủ sở hữu ###Owners equity	13273	13502	18155	25340	33208
27	Lợi ích của cổ đông thiểu số	0	0	0	0	0
28	TỔNG CỘNG NGUỒN VỐN ###TOTAL EQUITY	377324	380782	392850	389620	374323
29	KET QUA KINH DOANH	2013	Q4 2013	Q2 2013	Q1 2013	2012
30	Tổng doanh thu hoạt động kinh doanh###Gross Sale Revenues	433954	119051	110397	92099	363432
31	Các khoản giảm trừ doanh thu###Deduction revenues	1238	206	131	500	4293
32	Doanh thu thuần###Net Sales	432716	118845	110266	91599	359139
33	Giá vốn hàng bán###Cost of goods sold	385143	106304	96519	84727	321424
34	Lợi nhuận gộp###Gross profit	47574	12540	13747	6872	37715
35	Doanh thu hoạt động tài chính###Financial activities Revenues	26	1	4	3	35

From above data, we find financial ratios. After that, we encrypted data and divided it into five classes: very low, low, middle, high, very high.

DATA MINING

Step 1: Use Weka software to discover association rule of financial ratios in year 2013.
Step 2: Use Weka software to discover association rule of financial ratios in 3 years 2011 - 2013.

Research Findings

DISCOVERING ASSOCIATION RULES OF FINANCIAL RATIOS IN SAME YEAR

THE ASSOCIATION RULES OF FINANCIAL RATIOS IN SAME YEAR ARE LISTED IN FIGURE 4.

Conclusions 1: There is a significant positive relationship between accounts receivable turnover ratio, inventory turnover ratio, and total asset turnover.

Accounts receivable turnover ratio indicates the velocity of a company's debt collection, the number of times average receivables are turned over during a year. This ratio determines how quickly a company collects outstanding cash balances from its customers during an accounting period. It is an important indicator of a company's financial and operational performance and can be used to determine if a company is having difficulties collecting sales made on credit. Inventory turnover ratios is a measure of the number of

ISBN 978-602-19725-8-8

times inventory is sold or used in a given time period such as one year. It is a good indicator of inventory quality, efficient buying practices, and inventory management. This ratio is important because gross profit is earned each time inventory is turned over, also called stock turnover. Total asset turnover is a financial ratio that measures the efficiency of a company's use of its assets to product sales. It is a measure of how efficiently management is using the assets at its disposal to promote sales. The ratio helps to measure the productivity of a company's assets [17]. These ratios have a significant positive relationship.

Conclusions 2: There is a negative relationship between 3 ratios (accounts receivable turnover ratio, inventory turnover ratio, and total asset turnover) and profit margin ratio.

Furthermore, Napompech (2012) examines the effects of working capital management on profitability, from which it revealed a negative relationship between the gross operating profits and inventory conversion period and the receivables collection period. In other words, managers can increase the profitability of their firms by shortening the cash conversion cycle, inventory conversion period, and receivables collection period. Kulkanya Napompech (2012) showed a significant negative relationship between firm profitability and the inventory conversion period and receivables collection period. Mohammed Ziaur Rehman et al (2014) showed Total Assets Turnover Ratio has a negative relationship with Profit Margin and furthermore they showed that there has negative relationship between Total Assets Turnover Ratio (2014) and Net Profit Margin in Saudi Arabia. Deloof (2003) examined the correlation of Working Capital Management. The sample consisted of 1009 non-financial Belgian companies from 1992 to 1996. Results showed a significantly negative relationship between Gross Profits and the Average Period of Receivables. Samiloglu & Dermigunes (2008) in Turkey evaluated the effect of working capital on firm profitability. The purpose of the study was to consider consequences of relationship statistically between firm profitability and the constituents of cash conversion cycle. Sample consisted of Istanbul stock exchange listed manufacturing companies for period 1998 to 2007. Multiple regression models were used for analysis. The results of the study proved that inventory period, accounts receivable period and leverage negatively affect firm's profitability.

FIGURE 4: ASSOCIATION RULE OF FINANCIAL RATIOS IN SAME YEAR

1. Accountreceivable/sales_2013=very_low => Costofgoodsold/inventory_2013=very_low
2. Sales/Assets_2013=very_low => Costofgoodsold/inventory_2013=very_low
3. Costofgoodsold/inventory_2013=very_low => NetIncome/sales_2013 very_high
4. NetIncome/sales_2013 very_high => Accountreceivable/sales_2013=very_low Costofgoodsold/inventory_2013=very_low Sales/Assets_2013=very_low
5. Sales/Assets_2013=very_low => Accountreceivable/sales_2013=very_low
6. Accountreceivable/sales_2013=very_low => NetIncome/sales_2013 very_high
7. Sales/Assets_2013=very_low => NetIncome/sales_2013 very_high
8. Accountreceivable/sales_2013=very_low => Sales/Assets_2013=very_low
9. Costofgoodsold/inventory_2013=very_low => Sales/Assets_2013=very_low Accountreceivable/sales_2013=very_low
10. Sales/Assets_2013=very_low => Costofgoodsold/inventory_2013=very_low Accountreceivable/sales_2013=very_low

ISBN 978-602-19725-8-8

DISCOVERING ASSOCIATION RULE OF FINANCIAL RATIOS IN 3 YEARS

ASSOCIATION RULES OF FINANCIAL RATIOS IN 3 YEARS ARE LISTED IN FIGURE 5

Conclusions 3: If profit margin ratio is very high or inventory turnover ratio is very low or accounts receivable turnover ratio is very low in this year, profit margin ratio will be very high in next year

Profit margin ratio is very useful when comparing companies in similar industries. A higher profit margin indicates a more profitable company that has better control over its costs compared to its competitors. So if investors, enterprises can predict profit margin ratio in the future, they can make economic decisions better.

FIGURE 5: ASSOCIATION RULE OF FINANCIAL RATIOS IN 3 YEARS

1. Accountreceivable/sales_2013=very_low => Costofgoodsold/inventory_2013=very_low
2. Sales/Assets_2013=very_low => Costofgoodsold/inventory_2013=very_low
3. Costofgoodsold/inventory_2013=very_low => NetIncome/sales_2013 very_high
4. NetIncome/sales_2013 very_high => Accountreceivable/sales_2013=very_low Costofgoodsold/inventory_2013=very_low Sales/Assets_2013=very_low
5. Sales/Assets_2013=very_low => Accountreceivable/sales_2013=very_low
6. Accountreceivable/sales_2013=very_low => NetIncome/sales_2013 very_high
7. Sales/Assets_2013=very_low => NetIncome/sales_2013 very_high
8. Accountreceivable/sales_2013=very_low => Sales/Assets_2013=very_low
9. Costofgoodsold/inventory_2013=very_low => Sales/Assets_2013=very_low Accountreceivable/sales_2013=very_low
10. Sales/Assets_2013=very_low => Costofgoodsold/inventory_2013=very_low Accountreceivable/sales_2013=very_low

CONCLUSIONS

In this paper, we have studied the application of association rules to discover the relationship of financial ratios. We collected data from financial statements of real estate enterprises on the stock exchanges in Vietnam in the period 2011 to 2013. Preprocessing has been done to convert the data on financial statements to financial ratios. We used association rule mining method to extract association rules of financial ratios through Weka software.Results of this research help investors, banks, enterprises can predict other ratios in the future through association rules. It is useful for internal and external financial data users in making their economic decisions, including investing and performance evaluation decisions.

In the future, we are going to expand types of enterprises, financial ratios, and non-financial ratios.

ISBN 978-602-19725-8-8

REFERENCES

Bouckaert, R. R., Frank, E., Hall, M. A., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2010).WEKA---Experiences with a Java Open-Source Project.*The Journal of Machine Learning Research*, *11*, 2533-2541.

Chan, M. C., Leung, H. C., & Luo, W. D. (2005). Association mining system for financial ratios and stock prices in China and Hong Kong stock exchange.In *Soft Computing as Transdisciplinary Science and Technology* (pp. 1151-1160).Springer Berlin Heidelberg.

Deloof, M. (2003). Does working capital management affect profitability of Belgian firms?. *Journal of business finance & Accounting*, *30*(3-4), 573-588.

Han, J. (2001). KamberM.*Data mining: concepts and techniques*, *2*.

Ma, B. L. W. H. Y. (1998, August). Integrating classification and association rule mining.In *Proceedings of the 4th*.

Mahanavanont, T., & Jiang, J. A Study on the Working Capital Management of Thai Listing Firms.

Napompech, K. (2012). Effects of working capital management on the profitability of Thai listed firms. *International Journal of Trade, Economics and Finance*, *3*(3), 227-232.

Palepu, K., & Healy, P. (2007). *Business analysis and valuation: Using financial statements*. Cengage Learning.

Samiloglu, F., &Demirgunes, K. (2008). The effect of working capital management on firm profitability: Evidence from Turkey.

Sethi, A., & Mahajan, P. (2012). Association Rule Mining: A Review. *International Journal of Computer Science*, *1*(9).

Zhang, C., & Zhang, S. (2002). *Association rule mining: models and algorithms*. Springer-Verlag.

Zhang, D., & Zhou, L. (2004). Discovering golden nuggets: data mining in financial application.*Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, *34*(4), 513-522.

http://cophieu68.vn

http://www.cs.waikato.ac.nz/ml/weka/

[http://www.iasri.res.in/ebook/win_school_aa/notes/association_rule_mining.pdf

http://www.netmba.com/finance/financial/ratios/

http://www.readyratios.com/

ISBN 978-602-19725-8-8