

Analysis of Agrifood Purchasing Pattern Using Association Rule Mining¹

- Case of the Seoul · Gyeonggido · Incheon in South Korea -

Hyebin Jo · Young Chan Choe

Abstract Since the Free Trade Agreements (FTAs) with Chile, the EU, and the U.S., Korean agricultural produce markets have turned into a fierce competition landscape. Under these competitive circumstances, marketing is critical. The objective of the research presented herein is to understand the characteristics of customer preferences after locating trends of purchased items. So This research establishes sustainable strategies for Korean agricultural produce. This investigation used market-basket analysis techniques and panel data for its research. Market-basket analysis is a technique which attempts to find groups of items that are commonly found together. The results show that, for one year, processed food using wheat, processed marine products, and pork are commonly bought together and that yogurt and milk also are bought together. The characteristics of customers buying these items are 44 years old and live in a four-person household with two children. These customers do not live with their parents.

Keywords Association rule mining, Market basket analysis (MBA), agri · food product, purchasing behavior of South Korea.

1 Introduction

Due to the growth of global food enterprises and the conclusion Free Trade Agreements (FTAs) with Chile, EU, and U.S. , consumers are able to purchase a variety of inexpensive foods. Recently, with the rising interest in a healthy well-being, consumers purchasing foods have begun to look for eco-friendly and organic products. As the scale of the food market increases and the preferences of consumers become diversified, it is important for food enterprises to successfully identify the needs of customers.

Large enterprises can respond to changes in the market relatively quickly due to their rapid collection of market information and bargaining power in the distribution network. However, small- and mid-sized food enterprises and agricultural product producing organizations have difficulties in getting a competitive edge in the rapidly changing market environment because of the hardship in stable information about the market, and insufficient strategic marketing competence. Hence, it is important to provide products reflecting the needs of consumers at a proper point of time, and for this purpose, it is necessary to accurately identify the agrifood purchasing patterns of consumers.

The purpose of this study is to examine whether or not consumers have a specific purchasing pattern when buying agrifood. For example, identifying what products are simultaneously purchased by consumers purchasing milk. This study uses the market basket analysis (MBA) as the study method. The MBA is one of the data mining techniques also referred to as association-rule mining. The MBA has been used in various fields such as cross-selling, analysis of the traffic line of users on Websites, decision-making, credit evaluation in the financial sector, personal issues, etc.

This study involves the examination of 1,000 panel households residing in the capital area, and the examined

H. Jo · Y. C. Choe
College of Agriculture and Life Sciences
Seoul National University, South Korea

H. Jo(✉)
e-mail : jhb243@naver.com

Y. C. Choe
e-mail : aggi@snu.ac.kr

¹ This research was financially supported by the Rural Development Administration(RDA) of Korea (Grant NO. PJ008729)

data includes the details of agrifood purchases of the panels during the year of 2010. The MBA presents the association rule of the product purchase data of the panels; the association rule indicates which items are cross-purchased with each other. This study intends to draw the association rules by month and to examine the differences in the product mixes where cross-selling occurs between months and seasons. In addition, this study will examine the differences in the demographic characteristics of the customer base, illustrating a specific rule over several other association rules.

The results of this study will provide information on agrifood purchase patterns. These patterns can be utilized to establish marketing strategies for agrifood enterprises wishing to acquire competitiveness by rapidly responding to the needs of customers, including the development of bundled products.

2 Literature Review

In 2011, the scale of the Korean food and food service industry market was 133 trillion won, which is almost double of that of 10 years ago. The Korean food and food service market is expected to have a high probability of growth from this date on (Electronic Times Internet, June 13th, 2012). With the growth of the food industry in Korea, the propensity of consumers to consume certain foods is also changing. From this viewpoint, many studies have been conducted on food consumption in Korea.

Youngsu Jang and Kyounghee Lim (2000) found that in Korea, the food consumption has changed because of the swift improvement of income and the premiumization of food materials. Jang and Lim stated that the structural response in the aspect of production is very important through an analysis of food consumption. They conducted an analysis on consumption of foods, in particular, the consumption of fish and shellfish. They also conducted quantitative analysis and analyses using the dynamic demand function. The data used for their analysis was previously released statistical data related to food consumption from 1975 through 1997. As a result of quantitative analysis, urban households showed a gradual decrease in expenditures for traditional foods such as grain, marine products, and vegetables and an increase in expenditures for other items including meat. In view of the gradual increase in expenditures for processed products using fish and shellfish, it can be assumed that these expenditures are oriented to convenience. In addition, while intensive consumption was indicated for popular fish species until the mid-1980s,

the consumed fish species have become diversified since the late 1980s. As a result of the dynamic demand function, the researchers stated that the decrease in consumption of fish and shellfish should be interpreted as such: after the consumption of fish and shellfish reached its peak, the preference for fish species was changing from popular species to various minor species, rather than the decreasing consumption of fish and shellfish as replacement with meat.

Soyi Yoo (2005) measured the Berry index and Entropy index, which can explain the diversity of food consumption by using the data of food expenses of urban households. As a result of the two indices, values with opposite meanings were calculated, and it was concluded that there may be differences depending on the measuring method. As a result of analysis of the factors influencing the diversity of food consumption, all factors (household income, possession or non-possession of a house, residence in Seoul or other regions, presence or absence of a young child, number of employed persons in the household, and the number of people in the household) were significantly influencing the Berry index and Entropy index.

Jongseong Kim and Kyusu Hah (2010) intended to examine the general consumption realities of processed marine products and the factors influencing the consumer satisfaction and willing to purchase. As a result of correlation and multiple regression analysis based on the survey data, the product with a high purchase frequency of processed marine products was tuna (74.4%) with the average purchase frequency of 1 through 3 times per month (82.6%). The places of purchase were mostly large discount stores (71.7%). The satisfaction with purchase was high with 44.8% responding. Others responded ordinary (44.3%). As a result of the survey, which divided the factors influencing the willing to purchase into the product information, brand, recipe, and price, it was found that the product information was the most important factor.

Minah Kim et al. (2011) aimed to classify the types of consumers through cluster analysis and food consumption pattern by type by using the multinomial logit model. As a result of cluster analysis, consumers were classified into six types, and it was found that there was a difference in the consumption patterns between types. As a result of the multinomial logit model, it was found that factors such as household income, type of household, age of householder, possession or non-possession of a car, etc. influenced the classification of types of consumers.

Seungguk Choi (2011) analyzed the factors on consumer recognition and purchase decision of unheated fresh fruit juice and identified the purchase attributes on consumers of unheated fresh fruit juice. It was found that consumers

purchase products with fresh fruits for the health of themselves or their family. Consumers considering health as the reason for purchase of unheated fresh fruit juice represented 68.5% of the total respondents. This study examined the patterns of consumers from a marketing viewpoint with the demographic variables as the standard. As a result of analysis of the purchase or non-purchase, depending on the state of display in store, men had a higher concern of purchase than women. In addition, consumers under 30 years of age showed the highest concern of purchase. In relation to the purchase or non-purchase of product at taste events, the consumers with the educational level of graduation from college or higher had the highest desire for purchase, and the households with an income of 2 through 3 million won per month showed the highest desire for purchase. On the basis of these results, the marketing strategy for sales of fresh fruits was shown.

A number of studies have been performed on food consumption. However, the previous studies excluding the study of Minah Kim et al. (2011) and the study of Seungjuk Choi (2011) did not attempt to classify food consumption or to find the characteristics of consumers coming under a specific demographic. This study and the study of Minah Kim et al. (2011) intend to grasp the food consumption pattern. But this study has a difference with the study of Minah Kim et al., which is provision of the product mix with high association by using the MBA, and it then classifies the consumers purchasing the product mix. In addition, the study of Seungjuk Choi (2011) is similar to this study in that it aims to clarify the demographic characteristics of consumers purchasing a specific product, but it is different from this study in that this study deals not with a single product but with a product mix. Moreover, this study deserves attention in that it draws the product mix by conducting the MBA by month and by season.

3 Research Method

This study used the MBA in order to analyze the purchase data of consumers. The MBA is a useful technique in extracting the product bundle with association from the product purchase data and analyzing the purchase pattern of consumers (Chen et al., 2005). According to Agrawal et al. (1993), when there are two different products, X and Y, and a consumer purchases product X, one can examine if the consumer also purchases product Y; this is to clarify the purchase pattern through product association.

The MBA, which is introduced by Michael et al. (1997), consists largely of three concepts: support, confidence, and

lift. Support refers to the frequency of the purchase of product X and Y together for the total product trade frequency (N). Confidence, which is the concept to measure the accuracy of association rule, measures the degree of trade, including product Y of the trades of purchase of product X with probability. Finally, lift indicates the ratio of random purchase of product Y to the purchase of product X and Y together of the total purchases. The composing concepts can be shown as follows:

$$\text{Support} = \frac{n(X \mid Y)}{N} \quad \text{Expression (1)}$$

$$\text{Confidence} = P(Y \mid X) \quad \text{Expression (2)}$$

$$\text{Lift} = \frac{P(X \mid Y)}{P(Y)} = \frac{P(X \cap Y)}{P(X) \cdot P(Y)} \quad \text{Expression (3)}$$

According to the results of experiment on accuracy of the MBA, it was found that the accuracy is higher in the order of support, confidence, and lift (Namkyu Kim, 2008). That is, in order to be a meaningful association rule, the values of support and confidence should reach some level, and if these values are high, the value of lift should also be considered. However, the standard for the values of these indicators are not absolute, and thus the analyzer should repeatedly change the values by his or her judgment to find a meaningful rule. This study aims to draw the association rule by month by using the purchase data for one year and deduce a meaningful rule from these rules of support, confidence, and lift.

4 Data Collection and Sample Characteristics

The data used for this study is based on the panel database constructed by the Rural Development Administration as 'the agrifood consumption trend' business. For the operation of panels, 1,000 households were recruited from 200 Eups, Myeons, and Dongs (names of districts) in the capital area (Seoul, Gyeonggi-do, Incheon) in October 2009. However, only 995 households excluding the 5 households with a missing value were used for the analysis. The households attached the receipts for purchase of foods or register by day from January 2010 through December 2010; the receipts and registers were collected once a month. The information on agrifood was converted into data including: place of purchase, standard, price, indication or non-indication of cropping system, country of origin, etc. Agrifood was classified into some groups on the basis of the national food composition table as follows:

Table 1 Classification of Agrifood Products

Group	Contents
Processed Food	Snacks, kimchi, processed food using wheat, side dishes, sauces, processed food using rice, processed food using soybeans, tea, alcoholic beverages, instant food, HMR, beverages, seasoning, etc.
Grain	Sweet potato, rice, corn, bean, polished barley, etc.
Marine Products	Processed marine products, fish and shellfish, seaweed, etc.
Vegetables	Eggplant, potato, chili, root vegetables, herbs, garlic, white radish, Chinese cabbage, etc.
Dairy Products	Lactic acid (also known as milk acid), milk, other dairy products, cheese, etc.
Meat	Egg, pork, beef, chicken, etc.
Fruits	Banana, pear, peach, blueberry, apple, orange, plum, etc.

This study examined household size, income level, age of panels, number of children, etc. related to food purchase patterns. The average household size was three to five persons, and the majority (73.6%) did not live with their parents. The number of children was on average one to two persons at a rate of 74%. The income level of less

than 4 million won accounted for 55% of total respondents, and the age of panels were, on average, in their 40s at a rate of 40.1% of total respondents. The frequency of eating out was on average two to three times a month at a rate of 76%.

Table 2 Demographic Characteristics of Subjects of Survey

Category	Item	Number of samples (Household)	Ratio (%)	Category	Item	Number of samples (Household)	Ratio (%)
Household size	1 person	94	9.4	Income level	Less than 2 million won	148	14.9
	2 persons	72	7.2		2 through 2.49 million won	131	13.2
	3 persons	208	20.9		3 through 3.49 million won	156	15.7
	4 persons	456	45.8		3.5 through 3.99 million won	102	10.3
	5 persons	109	11.0		4 through 4.49 million won	88	8.8
	6 persons	37	3.7		4.5 through 4.99 million won	66	6.6
	7 persons	5	0.5		5 through 5.99 million won	66	6.6
	8 persons	3	0.3		6 through 6.99 million won	31	3.1
	9 persons	1	0.1		7 million won or more	39	3.9
	No response	10	1.0		No response	19	1.9

Number of children	0 person	49	4.9	Parents	No	732	73.6
	1 person	206	20.7		With both	34	3.4
	2 persons	530	53.3		With mother	42	4.2
	3 persons	95	9.5		With father	74	7.4
	4 persons	12	1.2		No response	113	11.4
	5 persons	3	0.3		-	-	-
	No response	100	10.1		-	-	-
Eat out	Less than below	52	5.2	Age	20-29	34	3.4
	Once/6 months	1	0.1		30-39	276	27.7
	Once/2 months	49	4.9		40-49	399	40.1
	Once/3 months	32	3.2		50-59	233	23.4
	Once/a month	186	18.7		60-69	37	3.7
	Twice/a month	296	39.7		No response	16	1.6
	Thrice/a month	271	37.2		-	-	-
	No response	108	10.9		-	-	-

5 Analysis and Results

This study used SAS 9.3 to analyze the product association rule. The product association rule had the frequency of purchase by panels as the standard. As a result of the analysis, a major rule was drawn, as shown in the following table, which reflects the dietary life of the panels.

The most frequently discovered association in the market basket during the year 2010 was that of processed marine products, pork, and processed food using wheat (Rule 11) and that of milk and yogurt (Rule 15). Product mix

with high lift, which means the nature of purchasing a product leads to the increasing possibility of purchasing other products, was analyzed to be that of milk-processed food using wheat- yogurt and that of snacks-milk-processed food using wheat. When excluding processed foods, the possibility to purchase together was high for the mix of pork and lettuce and that of bean sprouts and processed food using soybeans. It was analyzed that the association rule of processed foods was higher than that of meat, vegetables, etc. The 15 association rules frequently discovered in the market basket are provided in the table below.

Table 3 Major Association Rules in the Market Basket

Rule 1	Rule 2	Rule 3	Rule 4
Lettuce, pork	Pork, beef	Processed food using soybeans, soybean sprouts	Processed food using wheat, milk, egg
Rule 5	Rule 6	Rule 7	Rule 8
Processed food using wheat, milk, pork	Processed food using wheat, processed food using soybeans, milk	Processed food using wheat, processed food using soybeans, pork	Processed food using wheat, processed food using soybeans, processed marine products
Rule 9	Rule 10	Rule 11	Rule 12
Processed food using wheat, processed marine products, egg	Processed food using wheat, processed marine products, milk	Processed food using wheat, processed marine products, pork	Processed food using wheat, processed marine products, seashell
Rule 13	Rule 14	Rule 15	-
Processed food using wheat, yogurt, milk	White radish, spring onion	Yogurt, milk	-

This study drew the representative rules from No. 15 to analyze how different the rules were by month. The products coming under Rule 11 (processed food using wheat, processed marine products, pork) and Rule 15 (yogurt, milk) appeared almost every month. The rate of Rule 3 (processed food using soybeans, soybean sprouts) and

Rule 13 (processed food using wheat, yogurt, milk) being contained in the market basket together was high during six months from March through August. In November, Rule 14 including white radish and leaf mustard was higher than average.

Table 4 Association Rule by Month

Rule	L	S	C	Rule	L	S	C	Rule	L	S	C
Mar.				Apr.				May.			
Rule 1	2.66	2.75	47.32	Rule 1	2.36	2.54	40.9	Rule 1	2.33	2.74	41.42
Rule 8	2.19	2.72	38.75	Rule 8	2.31	2.86	41.75	Rule 3	2.19	2.45	33
Rule 11	2.33	3.22	41.09	Rule 11	2.5	3.68	45.1	Rule 11	2.38	3.25	40.66
Rule 13	2.84	2.4	51.37	Rule 13	2.54	2.35	43.46	Rule 13	2.62	2.15	43.76
Rule 15	2.48	3.81	44.93	Rule 15	2.34	3.67	40.08	Rule 15	2.51	3.71	41.85
Jun.				Jul.				Aug.			
Rule 5	2	2.1	33.65	Rule 6	2.03	2	37.16	Rule 3	2.41	2.41	34.06
Rule 11	2.37	2.02	33.37	Rule 11	2.47	2.46	36.16	Rule 8	2.27	2.28	38.1
Rule 13	2.85	2.1	47.94	Rule 13	2.49	2.22	45.48	Rule 11	2.48	3.29	41.56
Rule 15	2.51	3.53	42.24	Rule 15	2.28	3.7	41.7	Rule 13	2.29	2.05	43.05
Sep.				Oct.				Nov.			
Rule 2	2.21	2.16	32.41	Rule 1	2.58	2.28	43.74	Rule 3	2.35	2.29	32.1
Rule 3	2.36	2.48	33.02	Rule 3	2.27	2.83	35.69	Rule 10	2.1	2.29	32.04
Rule 6	2.07	2.03	35.88	Rule 7	2.22	2.46	37.69	Rule 11	2.47	2.44	36.95
Rule 11	2.55	2.72	41.03	Rule 11	2.4	3.1	40.99	Rule 14	4.54	3.16	38.49
Rule 15	2.25	2.98	38.99	Rule 15	2.32	3.16	39.25	Rule 15	2.68	2.98	41.06
Dec.				Jan.				Feb.			
Rule 3	2.14	2.55	32.1	Rule 1	2.54	2.08	52.93	Rule 8	2.08	2.91	39.79
Rule 7	2.04	2.27	34.11	Rule 3	2.11	4.3	39.74	Rule 9	2.25	2.01	42.98
Rule 8	2.07	2.45	36.88	Rule 4	2.21	2.09	39.86	Rule 11	2.18	3.38	41.6
Rule 11	2.17	3.11	36.28	Rule 13	2.54	2.45	45.83	Rule 12	2.37	2.15	45.38
Rule 15	2.36	2.92	37.24	Rule 15	2.36	3.65	42.51	Rule 13	2.71	2.17	45.48

This study analyzed the demographic characteristics of the consumers who put Rule 11 (processed food using wheat, pork, marine products) and Rule 15 (milk, yogurt) showed a high association throughout the year in their market baskets. During one year, the consumers who put Rule 11 in their market baskets were 854 persons, which account for 86% of the total panel of consumers, while those who

put Rule 15 in their market baskets were 808 persons, which account for 81% of the total panel consumers. It can be viewed that most panel consumers purchase the product mix of Rule 11 and Rule 15 throughout the year. The panels purchasing Rule 11 and Rule 15 have two children or eat out once to twice a month. In addition, 40% or more of them are in their 40s.

Table 5 The Result of Demographics Related to Rule 11 and Rule 15

Rule 11			Rule 15		
Household size	Frequency	Percent	Household size	Frequency	Percent
Missing value	6	0.7	Missing value	5	0.6
1	40	4.7	1	39	4.8
2	54	6.3	2	55	6.8
3	182	21.3	3	167	20.7
4	423	49.6	4	408	50.5
5	102	12	5	94	11.6
6-9	46	5.4	6-9	40	4.9
Sum	853	100	Sum	808	100
Number of children	Frequency	Percent	Number of children	Frequency	Percent
Missing value	44	5.2	Missing value	41	5.1
0	33	3.9	0	35	4.3
1	178	20.9	1	170	21
2	489	57.3	2	466	57.7
3	94	11	3	84	10.4
4	12	1.4	4	11	1.4
5	3	0.4	5	1	0.1
Sum	853	100	Sum	808	100
Parents	Frequency	Percent	Parents	Frequency	Percent
Missing value	46	5.4	Missing value	50	6.2
With both	31	3.6	With both	29	3.6
With either one	100	11.7	With either one	92	11.4
No	676	79.2	No	637	78.8
Sum	853	100	Sum	808	100
Eat out	Frequency	Percent	Eat out	Frequency	Percent
Missing value	45	5.3	Missing value	45	5.6
Less than below	42	4.9	Less than below	44	5.4
Once/2months	46	5.4	Once / 2months	41	5.1
Once /3months	26	3	Once / 3months	25	3.1
Once /a month	167	19.6	Once / a month	149	18.4
Twice/a month	270	31.7	Twice / a month	261	32.3
Thrice/a month	257	30.1	Thrice / a month	243	30.1
Sum	853	100	Sum	808	100
Age	Frequency	Percent	Age	Frequency	Percent
Missing value	7	0.8	Missing value	6	0.7
20-29	17	2	20-29	19	2.4
30-39	238	27.9	30-39	224	27.7
40-49	376	44.1	40-49	343	42.5
50-59	187	21.9	50-59	187	23.1
60-69	28	3.3	60-69	29	3.6
Sum	853	100	Sum	808	100

6 Discussion

The purpose of this study is to draw a specific cross purchase pattern in the agrifood consumption and to provide implications for the establishment of an agrifood marketing strategy. As a result of the analysis, many product mixes of processed products, milk, processed foods for meal replacement, etc. were discovered, and it is considered that this reflects Korean dietary life changing into a westernized taste. It should be noted that the results of this analysis do not vary greatly by season or by month. This seems to have been caused by the fact that processed foods account for a great part of foods and that vegetables, meat, marine products, etc. are supplied throughout the year.

The association rules of note purchased throughout the year are Rule 11, which is the mix of processed food using wheat, processed marine products, and pork, and Rule 15, which is the mix of milk and yogurt. This means, on the basis of Rule 15, that panels have a high probability to purchase milk and yogurt together. The type of consumers purchasing in this pattern were those in their 40s with two children and four household members. And they do not living with their parents. Therefore, when establishing the marketing strategy such as the development and release of new products, joint marketing, display at stores, bundling, etc., these consumer types need to be utilized. For example, marts can establish a plan referring to Rule 15 to sell the bundled products of milk and lactic acid (acid milk) that target the housewives with children.

According to the data of Statistics Korea in 2011, the population of the capital area (Seoul, Gyeonggi-do, Incheon) accounts for 49.5% of the total population of South Korea, and it was judged that the analysis of only the panels residing in this region would be significant. However, there are limitations in generalizing the result into the characteristics of nationwide consumers. Therefore, this study is proper as a subject to be studied after constructing a database related to national food consumption afterwards.

According to the 'Realities and Characteristics of One Person Households in 2010 Population and Housing Census' released by Statistics Korea, the ratio of one person households to general households is 23.9% (4.142 million households), which means one out of four households earns a living independently. This is a figure that increased by 8.4% from ten years ago (Electronic Times Internet* 2012). The ratio of one person households is gradually increasing in the food consumption market. Therefore it is

an important research task to analyze their characteristics in order to grasp a sales strategy reflecting single person household purchasing patterns.

References

- Minah Kim, Kwansu Kim, Donghwan Ahn (2011), Analysis of Causal Factors of Food Consumption Pattern, Journal of Korean Food Marketing Association, 28 (3)
- Jongseong Kim, Kyusu Hah (2010), Selection Attributes and Pursued Benefits of Processed Marine Products, Journal of Korean Society of Food Culture, 25 (5)
- Seungwoo Paik, *A Study on Operation Activation Plan for Agri-food Joint Marketing of Mountains Area* (13:3) 2009, pp 37-44.
- Soyi Yoo (2005), Diversity of Food Consumption of Urban Households, Consumption Culture Research, 8 (2)
- Jeongbin Lim, Donghwan Ahn, (2007). *Impacts of Korea-US FTA on Regional Agriculture of Gyeonggi-Do and Policy Implications* (9:3), pp 49-70.
- Youngsu Jang, Kyounghee Lim (2000), Characteristics of Consumption of Marine Products : Focused on the Urban Household, Food Distribution Research, 17 (3)
- Electronic Times Internet (The Aju Business/Reporter Wun Jeon), 2012.06.13., Korean Food and Food Service Market Doubled Than Ten Years Ago ... at '133 Trillion Won' Last Year
- Electronic Times Internet*** (Ulsan Maeil/Reporter Kyoungjin Kim), 2012.12.11., Era of 'One Person Household', Urgent Need for Tailored Policies
- Seungguk Choi, Purchase Attributes of Consumers on Unheated Fresh Fruit Juice, Master's Thesis at Konkuk University, Korea Education and Research Information Service, 2011
- Statistics Korea (Korean Statistical Information Service) (2010) http://kosis.kr/abroad/abroad_01List.jsp?parentId=A
- Ahn, K. I. (2012). "Effective product assignment based on association rule mining in retail," *Expert Systems with Applications*
- Chen, Y. L., Tang, K., Shen, R. J., and Hu, Y. H. (2005). "Market basket analysis in a multiple store environment," *Decision support systems* (40:2), pp 339-354.
- Cil, I. (2012) "Consumption universes based supermarket layout through association rule mining and multidimensional scaling," *Expert Systems with Applications*.
- Ford, P. B., and Dziewaltowski, D. A. (2008). "Disparities in obesity prevalence due to variation in the retail food environment: three testable hypotheses," *Nutrition reviews* (66:4), pp 216-228.
- Han, J., and Fu, Y. (1999). "Mining multiple-level association rules in large databases," *Knowledge and Data Engineering, IEEE Transactions on* (11:5), pp 798-805.
- Kim, N. (2008). "Effect of Market Basket Size on the Accuracy of Association Rule Measures," *Asia Pacific Journal of Information Systems* (18:2), pp 95-114
- Michael, J., and Gordon, S. L. (1997). "Data mining techniques: for marketing, sales, and customer support," *New York: John Wiley&Sons INC* (445).