Proceedings of the World Congress on Engineering 2013 Vol III,
WCE 2013, July 3 - 5, 2013, London, U.K.

The Effect of Clustering in the Apriori Data Mining Algorithm: A Case Study

Nergis Yılmaz and Gülfem Işıklar Alptekin

Abstract— **Many organizations collect and store data about their customers, suppliers and business partners. However, much of the useful marketing insights are hidden in that enormous amount of data. Data mining is the process of searching and analyzing data in order to find potentially useful information. Although data mining consists of a broad family of computational methods and algorithms, for this study, we have chosen the Apriori algorithm as the basis of the data analysis framework. The objective of the paper is to present the effect of clustering the data onto the association rules. Hence, we have compared the results of two different approaches: Finding association rules without consumer segmentation, and with consumer segmentation. The data analysis framework is applied to the data of mobile operating systems' users. By extracting most important information from consumer data, we claim that this framework directs providers offer the right product/advertisement to the right consumer.**

Index Terms— **Data mining, Apriori algorithm, K-means algorithm, clustering, user-oriented marketing strategy**

I. Introduction

The rapid growth of the smart phone applications market has fundamentally changed the way in which people access and consume content. This has contributed to a shift in competitive dynamics that impact network operators, operating system (OS)/ application store developers and handset manufacturers. According to Gartner Inc, free applications will account for 89% of total downloads in 2012. Worldwide mobile application store downloads will surpass 45.6 billion in 2012, with paid-for downloads totaling $5 billion [1]. In 2016, total global mobile application revenue is expected to reach $46 billion according to ABI Research. In this context, consumer's tendency to smart phone and accordingly to mobile OS market becomes more and more important. A fundamental truth in the business is that competitors have always an eye to others' customers. Besides, many customers are on the lookout for better services. It is known that acquiring new customers is more expensive than retaining them, an efficient customer retention strategy is crucial to a company's success [2].

This study focuses on finding behavioral patterns of consumers of mobile OSs. We believe that such knowledge would not only direct firms in the developing process of their own mobile applications, but would also contribute to mobile application marketing and advertisement. Data mining, in other words Knowledge Discovery in Databases, is an interdisciplinary research area that has significant impact both in academic and commercial environments. Data mining is processing data using sophisticated data search algorithms and statistical approaches in order to discover new meanings. It provides with extracting useful information from large volumes of data.

The data analysis framework in this paper is based on Apriori data mining algorithm [3]. Data on mobile OS consumers are collected through a questionnaire that is designed specifically for this purpose. We first apply Apriori algorithm to this data and obtain related association rules. An association rule expresses an association between items or sets of items. We utilize two metrics, *confidence* and *lift*, for evaluating these association rules. As a second approach, we again apply Apriori algorithm, but after clustering consumers. The segmentation of the consumers is generated using K-means algorithm [3]. The application is carried out on the WEKA package [4]. We compare the final results of these two approaches in order to reveal the impact of consumer clustering on the results of the Apriori algorithm.

The paper is structured as follows. Section 2 describes related literature. The methodologies and tools used in the framework are given in Section 3, while Section 4 presents detailed explanation of the proposed model. Section 6 reveals the results and the discussion is made in Section 7. Concluding remarks and future works are given in Section 8.

II. Related Work

It is possible to use different types of algorithms to extract most important information from a database. Ten most popular data mining algorithms identified by the IEEE International Conference on Data Mining (ICDM) are presented in [5]. They are listed as: C4.5, K-Means, SVM, Apriori, EM, PageRank, AdaBoost, kNN, Naive Bayes, and CART. These top ten algorithms are among the most influential data mining algorithms in the research community. Detailed explanation of most of these algorithms can be found in [3].

The tools and technologies of data warehousing, data mining, and other customer relationship management (CRM) techniques afford new opportunities for businesses to act on the concepts of relationship marketing [6]. In [7], the authors propose a comprehensive CRM strategy framework by segmenting the customers. They present a procedure to provide different kinds of usage analysis, including inter-cluster analysis and intra-cluster analysis.

Nergis Yılmaz, Institute of Science, Galatasaray University, nergisyilmaz87@gmail.com, Tel:+905057696615.
Gülfem Işıklar Alptekin, Department of Computer Engineering, Galatasaray University, gisiklar@gsu.edu.tr, Tel:+902122274480 Fax:+902122595554.

ISBN: 978-988-19252-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

III. Methodologies and Tools

A. Apriori Algorithm

The aim of the algorithms that determine association rules is to reveal relationships among variables that occur synchronously in large databases. An example of this type of algorithm is the market basket analysis [3]. If people who buy item *X* also buy item *Y* as well, we can state that there exists a relationship between item *X* and item *Y*. An association rule is an implication of the form $X \rightarrow Y$, where *X* is the antecedent and *Y* is the consequent of the rule. The association rules are defined as follows [3]:

Confidence is the conditional probability, $P(Y|X)$, which is what we normally calculate. Support shows the statistical significance of the rule, whereas confidence shows the strength of the rule. The minimum support and confidence values are set by company, and all rules with higher support and confidence are searched for in the database. If *X* and *Y* are independent, then it is expected lift to be close to 1; if the ratio differs – if $P(Y|X)$, and $P(Y)$ are different- it is expected there to be a dependency between the two items: If the lift is more than 1, it is seen that *X* makes *Y* more likely, and if the lift is than 1, having *X* makes *Y* less likely.

Support of the association rule $X \rightarrow Y$:

$$Support\ (X, Y) = P(X, Y) = \frac{\#\{\text{Customers who bought } X \text{ and } Y\}}{\#\{\text{Customers}\}} \quad (1)$$

Confidence of association rule $X \rightarrow Y$:

$$Confidence\ (X \rightarrow Y) \equiv P(Y|X) = \frac{P(X, Y)}{P(X)} = \frac{\#\{\text{Customers who bought } X \text{ and } Y\}}{\#\{\text{Customers who bought } X\}} \quad (2)$$

Lift, also known as interest of association rule $X \rightarrow Y$:

$$Lift\ (X \rightarrow Y) = \frac{P(X, Y)}{P(X)\ P(Y)} = \frac{P(Y|X)}{P(XY)} \quad (3)$$

Apriori algorithm [8] is interested in finding all such rules having high enough support and confidence, which has two steps:

1. Finding frequent item sets (those which have enough support), and
2. Converting them to rules with enough confidence by splitting the items into two, as items in the antecedent and items in the consequent.

B. K-means Algorithm

K-means is one of the simplest unsupervised learning and partitional clustering algorithms [9]. This algorithm classifies a given data set by finding a certain number of clusters (*K*). The clusters are differentiated by their centers. The best choice is to place them as much as possible far away from each other. The algorithm is highly sensitive to initial placement of the cluster centers. A disadvantage of K-means algorithm is that it can only detect compact, hyperspherical clusters that are well separated [10]. Another disadvantage is that due to its gradient descent nature, it often converges to a local minimum of the criterion function [11].

The algorithm is composed of the following steps (Fig. 1):

i. Initial value of centroids: Deciding *K* points randomly into the space which represent the clustered objects. These *K* points constitute the group of initial centroids.

ii. Objects-centroid distance: Calculating the distance of each object to each centroid, and assigning them to the closest cluster that is determined by a minimum distance measure. (A simple distance measure that is commonly used is Euclidean distance.)

iii Determine centroids: After all objects are assigned to a cluster, recalculating the positions of the *K* centroids.

vi. Object-centroid distance: Calculating the distances of each object to the new centroids and generating a distance matrix.

The whole process is carried out iteratively until the centroid values become constant.

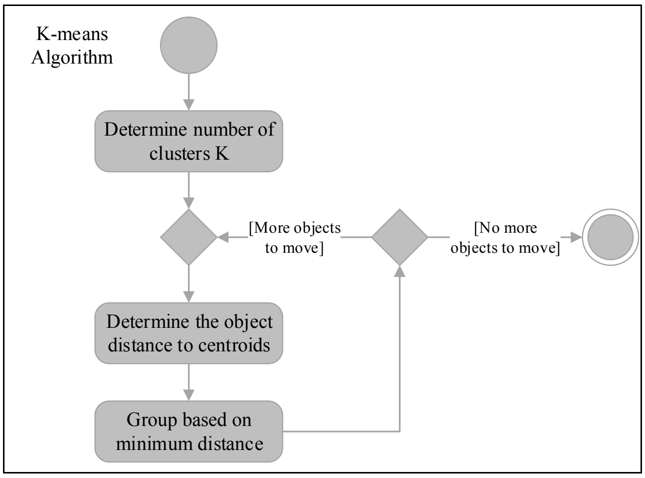

Fig. 1. Data and control flow of *K*-means algorithm

C. A Data Mining Tool: WEKA

WEKA is one of the popular software of machine learning that is written in Java [4]. It is free and available under the GNU license. The WEKA package supports several standard data mining tasks, more specifically, data preprocessing, clustering, classification, regression, visualization, and feature selection. Although it is possible to call WEKA from a Java code, we apply the Apriori and K-means algorithms directly to our dataset.

IV. Model

A. Research Framework

At the end of the case study, we aim at understanding consumer behavior and preferences when using mobile OSs on smart phones. Fig 2 illustrates the research framework. The first step involves developing the database. Doing so, we have observed the trends in mobile OS industry and have interviewed a marketing manager of Android OS during Google I/O. Using these knowledge, we have built the consumer questionnaire and accordingly, the tables in the relational database. The database is built on MySQL. The tables of the database reflect seven main information categories in the questionnaire. The answers coming from the consumers have been the input of the data mining process.

ISBN: 978-988-19252-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

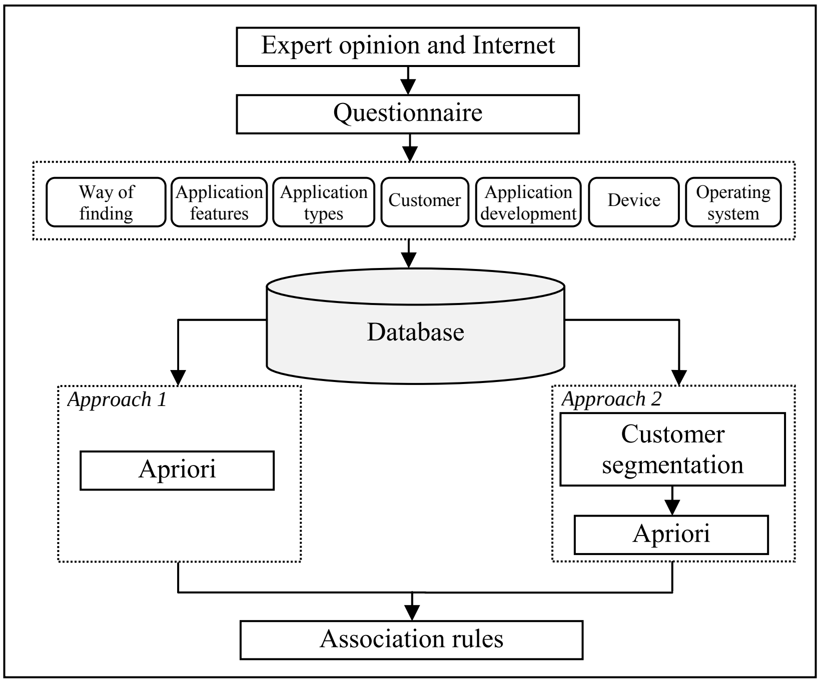

Fig. 2. Research framework

The main objective in this research is to show the effect of clustering to the extracting process of association rules. Therefore, we use two different approaches while mining the data. In the first approach, data are not been clustered and only the Apriori algorithm is applied to obtain related association rules. In the second approach, data are clustered using K-means algorithm, and then the Apriori algorithm is applied. The association rules are obtained from WEKA tool. The knowledge can then be integrated into the decision support process of telecommunication companies' marketing departments.

B. *Questionnaire*

For the case study, a field survey questionnaire approach is applied to collect information from consumers having mobile smart phones. The survey is answered by 209 consumers, where the number of male responders exceeds the number of female responders. The rate of female accounts for 43%, whereas the rate of male accounts for 57 %. Most of the respondents, 41.5%, are between 23–32 years old.

V. Results

At the time the questionnaire is given to consumers, the most common mobile OSs were Android and iOS in our country. According to all consumers' answers, both Android and iOS are found as more desirable by male consumers. Let us analyze in detail the results of two approaches separately.

A. Approach 1: Apriori algorithm without consumer clustering

Three rules are extracted from the demographic information, referred as R1, R2, and R3. The analysis results in Table 1 reveal that Android is found more attractive among male consumers, whereas iOS is more preferable by female consumers. Moreover, Android seems to be more desirable by the young consumers of age 23-32 years.

The association rules extracted from data are categorized into three main topic:

Table I.
Association rules from demographic information
(min sup: 20%; min Conf: 45%)

Rules	Support	Confidence	Lift	Consequent	Antecedent
R1	0.21	0.51	1.14	Next_OS=Android	AGE=23 - 32
R2	0.26	0.46	1.02	Nextt_OS=Android	Male
R3	0.20	0.47	1.01	Next_OS=iOS	Female

i. Knowledge pattern 1: Brand alliances: The minimum thresholds of support and confidence are set to 20% and 60%, respectively. The data analysis results in Table 2 show that two leaders in mobile OS market are Android and iOS, and their desirability is very close to each other. It seems that Android users want to continue to use Android as their next mobile OS. Similarly, iOS users choose iOS as their next OS. Therefore, brand loyalty has proven to be an important criterion that has influence on consumer choices, even in this small customer dataset. Telecommunication companies need to focus on attracting first time smart phone users.

Table II.
Association rules related to brand
(min sup: 20%; min Conf: 60%)

Rules	Support	Confidence	Lift	Consequent	Antecedent
R1	0.21	0.60	1.76	Current_OS =Android	Application Price Next_OS=Android
R2	0.27	0.60	1.71	Current_OS =Android	Next_OS=Android
R3	0.27	0.77	1.71	Next_OS =Android	Current_OS=Android
R4	0.30	0.78	1.67	Next_OS=iOS	Current_OS=iOS

ii. Knowledge pattern 2: Applications: Table 3 shows that the numbers of applications in the store and application prices have great effects on consumer's mobile OS preferences. Consumers who prefer non-paid applications will choose Android for their next OS. The data analysis results illustrate that application choices are shaped in respect to the gender and the age of the consumer. For instance, female users like social and communicational applications, whereas male users prefer transportation and

ISBN: 978-988-19252-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

game applications. Besides, users between ages of 23 and 32 like communicational applications. Since all the lift values are equal or greater than 1, we can state that the entire consumer group enjoys social and communicational applications.

iii. Knowledge pattern 3: Advertisement: The application advertising has huge expected revenues; hence, consumer behavior towards advertising is a critical information source for application providers. The results in Table 4 reveal from which channels consumers access to applications in applications store.

Similar to the application preferences, the chosen channel depends on the gender of the user. Female users reach the applications from social network advertisements, while male users find applications by searching the store. In general, most popular mobile applications are the social media applications, especially among young consumers. It would be profitable if the advertisement for young target group is placed into social media applications.

B. Approach 2: Apriori algorithm with consumer clustering using K-means

i. Cluster analysis with K-means algorithm: The clustering process which is generated by *K*-means algorithm reveals five meaningful groups of data. Obtained clusters are summarized in Table 5. The number of instances is the number of consumer data in the given cluster.

TABLE III.
ASSOCIATION RULES RELATED TO APPLICATIONS
(MIN SUP: 20%; MIN CONF: 60%)

Rules	Sup.	Conf.	Lift	Consequent	Antecedent
R1	0.24	0.69	1.36	Free Application	Next_OS=Android Application Price
R2	0.27	0.93	1.38	Free Application	Male Application Price
R3	0.25	0.71	1.40	Free Application	Current_OS=Android
R4	0.25	0.60	1.23	Application Type =Social	Female
R5	0.22	0.70	1.33	Male	App. Type = Transport
R6	0.28	0.64	1.17	App. Type= Communications	Female
R7	0.28	0.61	1.06	Male	App Type Game
R8	0.23	0.60	1.04	App. Type= Communications	AGE = 23 - 32

Let us interpret Table VI that represents the general view of clustering results. $Cluster_0$ is the most crowded group; therefore the highest cross selling opportunity is obtained. The most popular mobile OS is Android in this cluster. One of the most significant results is that, young population with relatively lower income prefers Android and plans to continue with Android. They are found as more sensitive to price, and they prefer Samsung as the mobile phone brand. $Cluster_0$ reveals cross selling opportunities in telecommunication market in Turkey. For instance Avea, a Turkish mobile operator, has a campaign for Samsung smart phones with Android, for the consumers of 23 to 32 years old.

In $Cluster_1$, female consumers generally use iPhone (iOS) and specify iPhone as their next choices. This cluster's mobile phone choices are influenced primarily from the brand, not from the price range. Hence, it would be a good idea to create social media campaigns for female consumers. While consumers of ages of 23 to 32 in $Cluster_0$ prefer Android, the same age group in $Cluster_1$ prefers iPhone (iOS) as their smart phone.

TABLE IV.
ASSOCIATION RULES RELATED TO ADVERTISEMENTS
(MIN SUP: 20%; MIN CONF: 75%)

Rules	Sup.	Conf.	Lift	Consequent	Antecedent
R1	0.20	0.79	1.15	Interesting Ads	Clicks on Ads Application Price
R2	0.38	0.89	1.45	Get App from Social Env.	Female
R3	0.43	0.75	1.16	Get App from Market	Male
R4	0.36	0.86	1.05	Get App from Social Env.	AGE = 23 – 32
R5	0.36	0.85	1.04	Get App from Social Env.	Get App from Market Male

TABLE V.
CONSUMER CLUSTERS AND THEIR NUMBER OF INSTANCES

	# of instances	Ratio (%)
$Cluster_0$	65	0.31
$Cluster_1$	41	0.20
$Cluster_2$	41	0.20
$Cluster_3$	42	0.20
$Cluster_4$	18	0.09

TABLE VI.
GENERAL CLUSTER MODEL

	Full Data	$Cluster_0$ (65)	$Cluster_1$ (41)	$Cluster_2$ (41)	$Cluster_3$ (42)	$Cluster_4$ (18)
Gender	Male	Male	Female	Male	Female	Male
Age avg.	23-32	23-32	15-22	23-32	15-22	23-32
Income (TL)	0-2000	0-2000	0-2000	2000- 4000	0-2000	2000-4000
Product Price	Medium	Medium	Low	Low	Medium	Low
Brand	Apple	Samsung	Apple	Apple	Blackberry	HTC
Current OS	IOS	Android	IOS	IOS	Blackberry	Android
Operator	Turkcell	Avea	Turkcell	Turkcell	Turkcell	Avea
Next OS	IOS	Android	IOS	IOS	Android	Android

In $Cluster_3$, Blackberry users are more crowded compared to iPhone users among university students. Majority of this cluster consists of female consumers who prefer Android as their next choices of mobile OS.

Opposite to $Cluster_0$, $Cluster_4$ includes consumers with relatively higher income value, but they are from the same age group. Consumers of $Cluster_4$ prefer HTC as the smart phone brand.

i. Knowledge pattern 1: Brand alliances: For the comparison purposes, we observe the values of the same association rules found in Approach 1. But actually, we find more rules with higher accuracy values using the Approach 2. Similar to the analysis of the Approach 1, Android and iOS are two mobile OS with the highest utilization rate and they are the most desirable ones (Table 7). The confidence and support values are higher than the ones in Approach 1, but the lift values are lower.

ISBN: 978-988-19252-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

ii. Knowledge pattern 2: Applications: The attributes *Next_OS*, *Current_OS*, *Brand* and the attributes related to consumer demographic information were dominating the association rules extracting from all data. Hence, we exclude them to observe in more detail the association rules related to applications and advertisements. The *K*-means algorithm is applied to this data subset, and four meaningful clusters are obtained (Table 8).

TABLE VII.
ASSOCIATION RULES USING CLUSTERING RELATED TO BRAND
(MIN SUP: 20%; MIN CONF: 60%)

Rules	Sup.	Confide.	Lift	Consequent	Antecedent
R1	0.5	0.70	1.05	Current_OS=Android	Application Price Next_OS= Android
R2	0.6	0.75	1.07	Current_OS=Android	Next_OS=Android
R3	0.6	0.87	1.07	Next_OS=Android	Current_OS=Android
R4	0.6	0.94	1.05	Next_OS=iOS	Current_OS=iOS

In $Cluster_0$, the male consumers of ages 23 to 32 usually prefer to find applications from application market and social media instead of Internet and media. 75% of the applications chosen from consumers of $Cluster_0$ are free, while 25% of them are paid. So, the advertisement in the applications would not reach $Customer_0$ group.

TABLE VIII.
CLUSTER MODEL FOR APPLICATIONS AND ADVERTISEMENTS DATA

Attribute	Full Data (207)	$Cluster_0$ (70)	$Cluster_1$ (51)	$Cluster_2$ (46)	$Cluster_3$ (40)
Age	23-32	23-32	33-50	15-22	23-32
Gender	Male	Male	Female	Male	Female
App. Price	Free	%75 Free - %25 Paid	%50 Free - %50 Paid	Free	Free
Market	Very High	High	Very High	Very High	Very Low
Social Environ.	Very High	Very High	Very High	High	Very High
Internet Ads	Low	Low	Medium	Medium	Very Low
App. Ads	Very Low	Very Low	Very Low	Low	Low
Media Ads	Very Low	Medium	Very Low	Very Low	Low

In $Cluster_1$, the majority of the users of paid applications are female of the age group of 35 to 50. Therefore, it is reasonable to attract more consumers by introducing free downloadable application campaigns.

The $Cluster_2$ consists of male consumers of age 15 to 22, who find their applications from the social environments, in other words from the friends. The advertisements that have this group as the target would be inefficient.

The $Cluster_3$ includes female consumers of age 23 to 32. Unlike other clusters, they do not find the applications by searching the application market. They mostly learn them in the social environment. Again, the media advertisement is not efficient for this group of consumer.

The reason that there are only three association rules in Approach 2 instead of eight is that, we select the same rules as in Approach 1, for the purpose of comparison (Table 9). After clustering the data, the confidence, the support and the lift values of R4, R6, and R7 all increase.

iii. Knowledge pattern 3: Advertisement: Both support and confidence values of three rules are higher than the support and the confidence values of the same rules in Approach 1 (Table 10).

TABLE IX.
ASSOCIATION RULES USING CLUSTERING RELATED TO APPLICATIONS
(MIN SUP: 20%; MIN CONF: 60%)

Rules	Sup.	Conf.	Lift	Consequent	Antecedent
R4	0.44	1	1.64	App type=Social	Female
R6	0.41	0.84	1.38	App type= Comm.	Female
R7	0.45	0.63	1.62	Male	App type= Game

TABLE X.
ASSOCIATION RULES USING CLUSTERING RELATED TO APPLICATIONS
(MIN SUP: 20%; MIN CONF: 75%)

Rules	Sup.	Conf.	Lift	Consequent	Antecedent
R2	0.39	0.78	1.15	Get App from Social Env	Female
R3	0.40	0.91	1.12	Get App from Market	Male
R4	0.34	0.75	2.02	Get App from Social Env	AGE=23 – 32

VI. DISCUSSION

We realized two different manner of data analysis on the same dataset. In the second approach, where we segmented the dataset into clusters, the analysis results are more reliable and consistent than the first approach, where the Apriori algorithm is applied directly to the dataset. The results are summarized in Fig. 5. The confidence values of all association rules in Approach 2 are higher than the ones in Approach 1, which show the strength of the rules. The reason is that, in the second approach we generate the data mining algorithm into a data subset which consists of more interrelated data. Similarly, the support values of all association rules found after clustering are higher than the ones in Approach 1, which show the statistical significance of the rules. K-means algorithm provides with grouping data with similar patterns. Therefore, we are able to find data that we miss in Approach 1, because in Approach 2, its statistical value (support) is higher.

VII. CONCLUSION

The intense competition and increased choices available for customers have created new pressures on marketing decision-makers and there has emerged a need to manage customers in a long-term relationship. If companies make sense of customer needs and manage the relationships more intelligently, it is obvious that they will provide crucial competitive differentiation to gain market share and retaining customers. Customer retention marketing is a tactically-driven approach based on customer behavior.

This study uses a research framework which can be appropriate for any sector to mine customer knowledge. The data mining is realized using two of the most known data mining algorithms: Apriori algorithm and K-means algorithm. They both help us to find association rules. We have compared the resulting association rules in two different data analysis approaches. In the first analysis, data are not clustered, whereas in the second analysis data are clustered.

ISBN: 978-988-19252-9-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

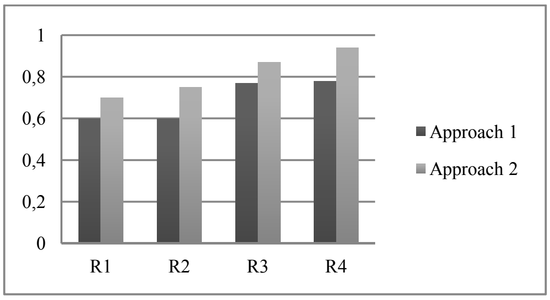

Fig. 5.a. Confidence of data related to brand alliance

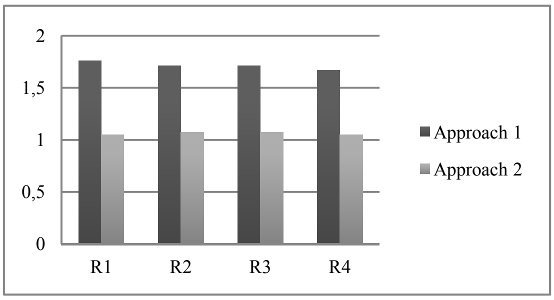

Fig. 5.b. Lift of data related to brand alliance

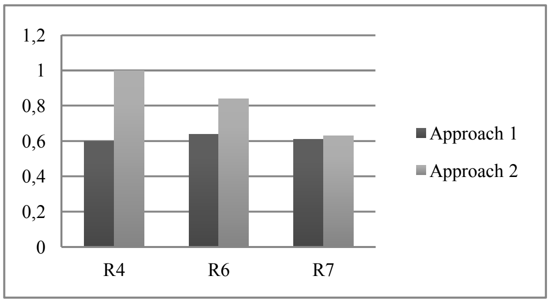

Fig. 5.c. Confidence of data related to applications

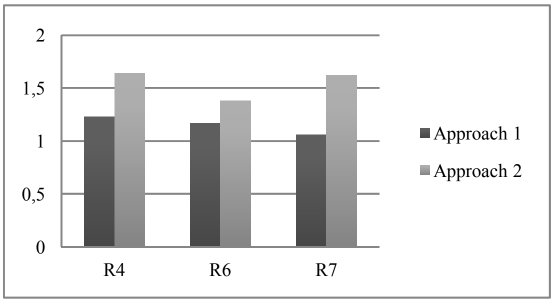

Fig. 5.d. Lift of data related to applications

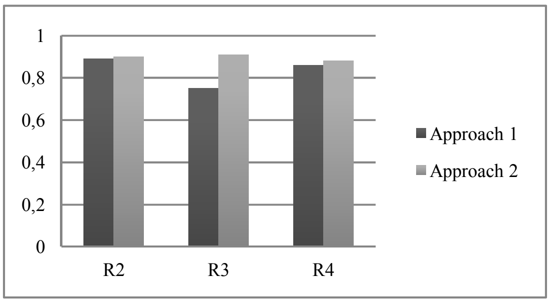

Fig. 5.e. Confidence of data related to advertisement

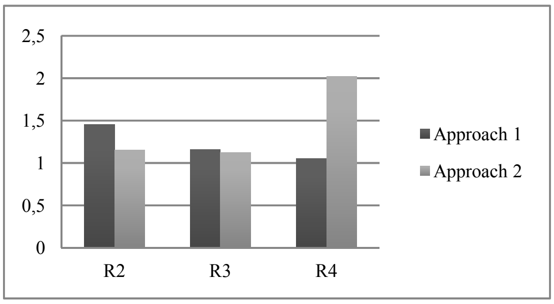

Fig. 5.f. Lift of data related to advertisement

Both approaches then use Apriori algorithm to extract related association rules. In this manner, we aim at determining the impact of clustering into our case study.

As the case study, we have chosen the mobile operating system industry in Turkey. It is doubtless that our customer knowledge of 209 users is not enough to extract general strategies for mobile OS industry, and it is not reasonable to state the results as certain especially in such a tremendously changing market. We have chosen this area for demonstrating the usefulness of the approach in a simple case. Consumer information is received through a questionnaire. The responses to this questionnaire were taken in mid-2012, and now, we have seen that the results are accurate.

In data mining, in order to get more reliable results, it is important to implement different algorithms and find the one with the best predictions. Going forward, we will implement more clustering algorithms and compare their results, especially for this kind of small datasets.

ACKNOWLEDGMENT

This research has been financially supported by Galatasaray University Research Fund with the project number 12.401.007.

REFERENCES

[1] C. Newark, Mobile Apps Put the Web in Their Rear-view Mirror, http://blog.flurry.com/bid/63907/Mobile-Apps-Put-the-Web-in-Their-Rear-view-Mirror

[2] E.W.T. Ngai, L. Xiu, D.C.K. Chau, "Application of data mining techniques in customer relationship management: A literature review and classification", *Expert Systems with Applications* 36, 2009, pp. 2592-2602.

[3] E. Alpaydın, Introduction to Machine Learning. The MIT Press, London, 2010.

[4] I.H. Witten, E. Frank,L. Trigg, M. Hall, G. Holmes, S.J. Cunningham, "Weka: Practical machine learning tools and techniques with Java implementations", in Proc: Workshop on Emerging Knowledge Engineering and Connectionist-Based Information Systems, 1999, pp. 192-196.

[5] X. Wu, V. Kumar, J.R. Quinlan, J.Ghosh, Q. Yang, H. Motoda, G.J. McLachlan, A. Ng, B. Liu, P.S. Yu, Z.-H. Zhou, M. Steinbach, D.J. Hand, D. Steinberg, "Top 10 algorithms in data mining", *Knowledge Information Systems* 14, 2008, pp. 1–37.

[6] C. Rygielski, J.C. Wang, D.C. Yen, "Data mining techniques for customer relationship management", *Technology in Society* 24, 2002, pp. 483-502.

[7] L.C. Cheng, L.M. Sun, "Exploring consumer adoption of new services by analyzing the behavior of 3G subscribers: An empirical case study", *Electronic Commerce Research and Applications* 11, 2012, pp. 89-100.

[8] R. Agrawal, R. Srikant, "Fast algorithms for mining association rules in large databases", in Proc: International Conference on Very Large Databases, 1994, pp. 487-499.

[9] A.K. Jain, M.N. Murty, P.J. Flynn, "Data clustering: A review", *ACM Computing Serveys* 31(3), 1999, pp. 264-323.

[10] M.E. Çelebi, H.A. Kingravi, P.A. Vela, "A comparative study of efficient initialization methods for the k-means clustering algorithm", *Expert Systems with Applications* 40, 2013, pp. 200-210.

[11] S.Z. Selim, M.A. Ismail, "K-means-type algorithms: A generalized convergence theorem and characterization of local optimality". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6(1), 1984, pp. 81-87.