Applying Data Mining Method for Marketing Purpose in Social Networks: Case of Tebyan

Abstract: Within a very short period of time, social networking sites are developed among different users all around the world. Social networks have high value to business intelligence. In these networks, there are so many advantages and demands on addressees and their interest recognition. How do we increase our social network users, posts, and effectiveness? How many consumers be segmented with respect to their reactions to social network? The creation of a target market strategy is integral to developing an effective business strategy. The purpose of this article is market segmentation and correctly identifying the target groups for social network using data mining techniques. As users in each segment have their own and specific interests, social networks can define them by their demographic profiles, they can also change their development strategies according to users and interests they want to engage in. In this research we deploy Data Mining methods for segmenting Tebyan social network users to see how this method could contribute toward marketing strategies and purposes. According to K-mean algorithm, we demonstrate 5 different customer categories based on their characteristics and behavior that deploying appropriate strategy for each category can help the marketing performance.

Keywords: Data Mining, Social Network, Marketing, Segmentation, Marketing Strategy, Electronic Marketing, Target Marketing, Two-Step, Kohnen, K-means

1. Introduction

Having the precise information is crucial for making the right decision. The problem of collecting and storing data, which used to be a major concern for most organizations, is almost resolved. Now, organizations will be competing in generating information and even knowledge from gathered data.

The internet has become mainstream in everyday communications and transactions. Research has traditionally linked demographic factors to web use. Mostafa (2006), for instance, found a positive impact for educational level on internet use, while age had a negative impact. Eastman and Iyer (2004) posit that age is an important factor in explaining the attitude toward actual use of the internet. Whereas, Gefen and Straub (1997) argue that gender affects the perceptions and meanings rather than the actual use of e-

mail, also some recent studies highlighted some behavioural gender differences (Mostafa, 2006; Lyndon Simkin, 2008).
The rapid rise in social media marketing is reflected in increased scholarly attention on this topic, as researchers increasingly exploring the marketing approaches and strategies now available through social media (Fournier and Avery, 2011; Hanna et al., 2011; Kaplan and Haenlein, 2011a; Kietzmann et al., 2011; Weinberg and Pehlivan, 2011; Stephen P. Borgatti et al, 2014). As traditional marketing communication is changing, we need to return to basics, and realize how consumers use and react to social networks.
Based on importance of data analysis in today's trades, especially social networks as recently emerging important filed in internet based businesses, our purpose in this study is to demonstrate the application of data mining to social networks users' segmentation. We focus on segmenting Social network user's behaviour and profiles. Then found how data mining could be useful for a right strategic decision making of social networks and its marketing purpose and how its techniques can yield good results in user attraction. We try to demonstrate importance of extracting meaningful information about users based on available data in social networks to enable effective marketing policies based on users, needs and specification.

1.1 Social Networks

Social media is defined as a group of internet-based applications that build on the ideological and technological foundations of Web 2.0, allowing for the creation and exchange of user-generated content (Kaplan and Haenlein, 2010). Within the contexts, social media applications exist to facilitate user interaction, and include blogs, content communities, discussion boards and chat rooms, product and/or service review sites, virtual worlds, and social networking sites (Kaplan and Haenlein, 2010; Mangold and Faulds, 2009). In this paper we focus on social networking, which refers to applications, such as Facebook, Twitter and Tebyan, that allows individuals to construct personal profiles and develop a list of others with whom they share and interact with (Boyd and Ellison, 2007; Stephen P. Borgatti et al, 2014).
Growth trends of social networking are seen all over the world. Social networking dominates consumer time spent online, garnering an average 54 minutes a month per person and growing fastest among users aged 55 and above (comScore, 2012a; Nielsen, 2011). It includes smartphone ownership as a covariate in line with extended view of social media and "smart" mobile phones as enabling tools for social networking (comScore, 2012b; Hanna et al., 2011) which bring together the social and mobile channels (Rheingold, 2002). Unlike traditional mobile phones, smart-phones are data-centric and

capable of running third-party software, such as social networking applications (Cheng et al., 2007; Stephen P. Borgatti et al, 2014).
Currently web users spend considerable time on social networking sites which is of great importance to them. Considering this day-by-day growth and importance of social networks, study of these sites on a business view is vitally important. The current usage of social networks as a huge market, makes these studies more and more important for today’s business.

1.2 Marketing, strategies and segmentation

The creation of a clear target market strategy is at the heart of an effective business strategy. Despite the acknowledged benefits of the concept of market segmentation, many organisations fail to adequately embrace the concept (Paola Carone and Simona Panaro, 2014). The marketing strategy will be implemented in a better way as effect of analysis of customer data.
Views of leading business executives as to the value of market segmentation show that many senior managers believe it matters. Segmentation is about knowing which customers are not worth targeting. Consumer segmentation permits companies to gain a strategic advantage over their competitors by helping them to identify the unique attitudes and needs of the divergent segments and thus to translate strategic opportunities into an actionable plan (Dibb et al., 2002; Lyndon Simkin, 2008). The creation of a target market strategy is integral to developing an effective business strategy. The concept of market segmentation is often cited as pivotal to establishing an effective target market strategy especially for today’s business.
Thus a review of the online consumer segmentation literature is first conducted. Consumer segmentation can be used to identify naturally occurring customer groups and to provide an understanding of each segment motives, characteristics, and needs (Swinyard, 1996). In fact, the potential benefits to be gained far outweigh the resource implications required to implement a successful segmentation approach (Quinn, 2009, p. 254; Lyndon Simkin, 2008). Segmentation enables an organization selectively to target “good” business (Anderson and Narus, 2003; Dibb et al., 2002). Weinstein (2004) comprehensively argues a persuasive case for practising market segmentation.
Hunt (1991, p. 176) underscores the importance of segmentation studies in marketing, indicating that:
Classified schemata play fundamental roles in the development of a discipline since they are the primary means for organizing phenomena into classes or groups that are amenable to systematic investigation and theory development. (Lyndon Simkin, 2008)

Consumer segmentation analysis is stated to allow the segments to be identified based on natural associations observed during data analysis typically via cluster analysis techniques (Wedel and Steenkamp, 1989; Swinyard, 1996; Lyndon Simkin, 2008).

One of the underlying reasons top managers fail to embrace best practice market segmentation is their inability to manage the transition from how target markets in an organisation are currently described to how they might look when it is based on customer characteristics, needs and decision-making. In creating market segments, this approach also develops a better understanding of customers, which is a rather crucial requisite for effective marketing (Vandermerwe, 2004). Even when market segments are identified, many organisations struggle to operationalize these segments (Craft, 2004; Dibb and Simkin, 2008; Hassan and Craft, 2005; Palmer and Millier, 2004; Paola Carone and Simona Panaro, 2014).

Segmentation is one tool that can be employed in understanding how consumers differ in terms of their interaction with and behavioural responses to social network marketing. Segmentation is a fundamental component of marketing and drives more precise targeting and positioning, which ultimately increase the creation of customer value (Wedel and Kamakura, 2002). Even through a non-strategic lens, segmentation can be a powerful descriptive tool (Wedel and Kamakura, 1999), particularly for investigating how groups of consumers behave. The knowledge gained from the tool can be used to measure return on investment of marketing campaigns and make better online business decisions. With respect of this decision organizations can improve revenues and reduce forced markdowns.

In the firms which the social network is the base of their activity, segmentation can be the fundamental part for their marketing and business strategy. While the use of social network impose its limitation to the firm, but in a virtual business accepting these limitations are inevitable. Anyhow an important segmentation tool can support the firm strategies and help the firm to progress toward its' marketing goal.

1.3 Data Mining

In the current post-industrial information society, data are one of the most valuable resources. However, data are not useful, because they have hidden information, unless they can be processed and turned into information that allows enterprises a competitive advantage. Terabytes of data are generated every day in many organizations. These data are a strategic resource. The technologies for generating and collecting data have been advancing rapidly. At the current stage, lack of data is no longer a problem; the inability to generate useful information from data is! The explosive growth in data and database results in the need to develop new technologies and tools to process data into useful information and knowledge intelligently and automatically (Ajay Mehra, 2014). While data may be in many disparate forms and formats, essentially making it unusable.

To extract hidden predictive information from large volumes of data, data mining (DM) techniques are needed. Organizations are starting to realize the importance of data mining in their strategic planning and successful application of DM techniques can be an enormous payoff for the organizations (Ajay Mehra, 2014). In general, traditional statistical models only can handle a limited data and test the existing hypotheses (Chopoorian et al., 2001). Traditional data analysis methods often involve manual work and interpretation of data that is slow, expensive and highly subjective (Fayyad et al., 1996; Groenewegen and Moser 2014). Thus, the KDD function in data mining can be used to face the challenge and increase business competitive advantages. On the one hand, data mining can easily handle large data. On the other, it facilitates both uncovering relationships hidden in complex data and identifies unknown problems and opportunities (Campbell et al, 2014).

Data mining is extracting or mining knowledge and valuable information from large volume of data (Han and Kamber, 2006; Groenewegen and Moser 2014). Data mining is an iterative process of searching for new, previously hidden information in large volumes of data (Kantardzic, 2003; Devedzic, 2001; Budeva and Mullen, 2014). It is the process of nontrivial extraction of implicit, previously unknown and potentially useful information such as knowledge rules, constraints, and regularities, using pattern recognition technologies as well as statistical and mathematical techniques (Mehra et al, 2014).

Data mining, fondly called patterns analysis on large sets of data, uses tools like association, clustering, segmentation and classification for helping better manipulation of the data (Groenewegen and Moser, 2014) since data mining techniques are very popular, many researchers have applied them in various domains (Budeva and Mullen, 2014). Considering enormous volume of data creation in digital age, interests and applications of data mining are increasing in many businesses (Campbell et al, 2014).

The KDD process is viewed from the broadest scope that incorporates wide-ranging activities:

- data collection;
- data cleaning;
- data analysis;
- scoring the database;

- decision-making support; and
- model refinement (Campbell et al, 2014).

The widespread application of the data mining techniques has resulted in many numbers of data mining tools in the market. These tools provide the GUI interface for better interactivity with the business organization managers. Ranjan and Bhatnagar (2008b) showed the effect of various data mining tools from the data mining perspectives. Data mining helps business world by providing two different forms of information:

- Descriptive. This form can be interpreted as discovery of unknown information from huge volume of data. The information is mined from customer data. This information helps to know the current status of the organization.
- Predictive. This form uses the descriptive information to predict future trends. This helps the business managers to take future strategic decision for growth and development in the organization (Aljukhadar and Senecal, 2011).

The prediction of unknown information helps business managers to plan the marketing strategy for the future. Data mining provides the in-depth analysis of the customer related data, which helps the organization to make crucial strategic decision (Aljukhadar and Senecal, 2011). Extracting useful information from data can be a complicated and sometimes a difficult process. Data mining tools are then used to uncover useful patterns and relationships from the data captured (Mehra et al, 2014). The powerful data mining algorithms help in analysis of huge volume of customer data which is the need of most organizations. The data mining is supported by various techniques like clustering, classification, prediction, association, genetic algorithms, and neural network.

Clustering/segmentation are fundamental data mining analysis tasks (Peacock, 1998). Although many statistical tools can perform cluster analysis, data mining provides numerous and irreplaceable benefits. First, the software is easy to use and unlike traditional statistical methods and tools, data mining does not require sophisticated statistical knowledge and data pre-processing (Chen and Sakaguchi, 2000). Thus, advanced statistics training is not necessary for managers or administrators to take advantage from this method. Second, data mining tools automatically provide more informative tables and visual charts, which present data analysis results from various perspectives to facilitate understanding. Hence, data mining techniques can better support decision-making processes. Third, data mining’s most attractive function is that it can be used as a knowledge discovery in database (KDD), which means that data mining is a process of

discovering hidden patterns in large consolidated databases (Campbell et al, 2014).
Our main objective is to use data mining techniques to recommend customer segments for our marketing target in social networks. Data mining here helps to examine consumer characteristics and behaviour in terms of prescription renewal and marketing.

2. Literature Review

Academics and practitioners claim that whereas a majority of consumers use the internet regularly, they use it for multiple, divergent purposes (Kau et al., 2003; Mathwick, 2001; Pew Internet and American Life Project, 2010). That is, consumers that use the internet do not seem to form one homogenous marketing group (Lyndon Simkin, 2008). Srinivasan et al. (2002) examined the antecedents and consequences of customer loyalty in an online context and identified the factors that significantly affect e-loyalty (Lyndon Simkin, 2008). Other researches indicated that online shoppers are most likely to be younger males with high internet experience, greater education, and higher income (Li et al., 1999; Sin and Tse, 2002; Swinyard and Smith, 2003; Lyndon Simkin, 2008).
Several studies have provided a valuable basic understanding of online consumer interactions through consumer segmentations on the basis of dimensions such as usage or motivations to participate (e.g. Foster et al., 2011; Ip andWagner, 2008; Li and Bernoff, 2008; Riegner, 2007; Wasko and Faraj, 2000a, b; Wiertz and DeRuyter, 2007; Borgatti, 2014). Previous researches have also asserted that a range of demographic variables can influence online behaviour, however, results are inconsistent (Konus et al., 2008). Commonly employed demographic variables in determining online consumer usage behavior include age, gender, education, and income (Kushwaha and Shankar, 2008; Strebel et al., 2004; Borgatti, 2014).
A study by Foster et al. (2011) integrates much of the previous studies on online social media behaviour by offering a segmentation analysis using several bases. Creating content, socializing with others, and seeking information are the three dimensions used in their survey about university students. Four segments emerge along two axes of behaviour – information needs and participation – with “mavens” being high on both these dimensions and the minimally involved low on both (Borgatti, 2014).
Moreover, data mining continues to attract more and more attention in the business and scientific communities. In a 1997 report, Stamford, Connecticut-based Gartner Group mentioned: “Data mining and artificial intelligence are at the top of the five key technology areas that will clearly have a major impact across a wide range of industries within the next three

to five years" (Mehra et al, 2014). As data mining has such powerful data analysis functions, it has been used in a variety of businesses (Campbell et al., 2014). Lejeun (2001) discussed the impact of applying data mining techniques on churn management. The data mining has found its application in many sectors of the business. Data mining has a direct impact on analytical results that will drive business decisions (Berson et al., 1999; Aljukhadar and Senecal, 2011).

Ahmed (2004) defined classification as the way to discover the characteristics of customers who are likely to churn and it also helps to predict who those likely customers are. Clustering algorithms like K-mean algorithms helps in segmentation process. In this algorithm various group of segment are made based on the characteristic; the similar characteristics are put in the same group. This technique helps to identify the potential customers for the organization. On the other hand, prediction technique contributes to planning the strategy for the future. Furthermore, associations ease identifying rules of affinities among the collected customer data. For example important marketing applications of the association rules include market basket analysis, attached mailing in direct marketing, fraud detection, etc. And finally the neural network can be used for predictive models for attrition, churn management and can also be deployed for calculating the customer life time value (Aljukhadar and Senecal, 2011).

The purpose of this paper is to propose a solution for social network managers, focusing on customer usage behaviour, which evolves from the organisation's existing criteria used for grouping its users. The study reveals, when compared with traditional statistical methods, that data mining provides an efficient and effective tool for market segmentation.

3. Introducing Tebyan Social Network

Tebyan portal was founded in 2001 and it has attracted about one million users so far. Tebyan services are available in 8 different languages which consist of Persian, Arabic, English, French, Russian, Turkish, Kurdish and Urdu. The main activity of Tebyan is focused on cultural affairs fitting to Islamic-Iranian culture with most users being young. According to Alexa website ranking, Tebyan is ranked the 28th in Iran.

Tebyan is a portal which consists of many different sites; Social Network is one of these sub-sites which has more than 500,000 users so far. Tebyan Social Network is modelled from Facebook and has most of its functionalities such as sharing, chatting, commenting, like, groups, etc.: and has 48 discussion forums. Also there is an online market on Tebyan, which can be used by users for purchasing or ordering goods.

Tebyan managers make full database accessible for authors in order to analyse based on current study purpose. Consequently, all related data are exported from the database which is including different information of users, their demographic data based on their profile, their usage history, etc. According to confidentiality rules of portal, private user information comprising the names, email addresses and etc. are excluded from the exported data. Afterwards, based on the followed methodology, raw data are filtered, pre-processed and prepared for analysis serving our purpose.

4. Methodology

The data panel comprised about 170,000 site members. Sample selection was randomly performed following an iterative process. A random iterative selection process was done to ensure a representative sample and made a sample consisting 10345 records. This step has been done by considerable delicacy, as respects that it is very important to pre-process the data before continuing with the data mining process (Larose, 2005), otherwise the end-results will be misleading (Budeva and Mullen, 2014).

The segmentation method employed in this research is describing in the following section. We consider demographic specifications of users to investigate their possible value in segmenting consumers. Gender, age, income, and educational level are the demographic variables in our analysis. A specific research summarizes functions, tools and software commonly used in data mining (Peacock, 1998; Cheng et al., 2005; SAS Institute Inc., 2002; Thelen et al., 2004). In the stated research are five fundamental functions identified:

(1) summarizing;
(2) classification;
(3) prediction;
(4) segmentation; and
(5) link analysis (Campbell et al, 2014).

We intend to segment users of Tebyan Social Network to determine target market. In the first step of data mining, we describe data collection and pre-process them. Then method and results of clustering will be explained later. First of all, the following data was extracted from Tebyan Social Network database.

- User's demographic profile: age, sex, education level, employment status and location
- The number of friends of each user
- The number of sent and received friend request
- Membership time in months

- The number of posts of each user
- The number of likes in each page on same category; (Tebyan consists of pages which have one or more categories).

In this social network database, each user is defined with a unique ID. All data are converted to numerical value and then they would be normalized. We have 10345 records in our data set.

One of the main concerns in pre-processing is missing values. Like other normal researches, we are faced with this problem in our data. There are different ways to fix this problem like deleting related records, replace them with a mean value, median or mode of other records, check all possible values in features and find the most similar record to our record. In this research we replace the missing values with the mean values of each feature.

Clementine is one of the best software applications in data mining which we use for clustering of our prepared dataset. Clementine has implemented some clustering algorithms. We used K-means, Kohonen, SOM and Tow-Step algorithms for our purpose. While our main goal is to segment the target market, the property of membership time, the number of posts, the number of friends and sent and received friend request is filtered by filter tools in Clementine.

5. Findings

It should be noted in used dataset, categorising is one of the most important features to distinguish clusters, and so in the output of the clustering, if the value of this feature in different clusters are the same, the clustering was not done properly. Also educational level and job can be used as a reference for a decision. It is better that output clusters have a normal distribution. In this way and according to the mentioned criteria, we represent and check the output of these algorithms for different parameters.

5.1. Two-Step Algorithm Result

Data is distributed in 6 clusters with this algorithm. Data distribution based on this algorithm is shown as below in table 1:

Title

Cluster #	Number of Records (%)	Age (avr.)	Favourite Category		Number of likes in this category (Median)	Education		Job		Marital Status (Single: %)	Sex (Male: %)
			Median	Mode		Median	Mode	Median	Mode		
C#1	15.66	32.0	Science	Religion	4	Higher National Diploma	BA & Bsc	Industrialist	Labour	19	66
C#2	4.61	40.8	Culture	Comedy	6	Higher National Diploma	Bsc & BA	Student	Student	7	60
C#3	15.16	19.5	Sport	Comedy	5	High School Diploma	High School Diploma	Student	Student	86	62
C#4	14.93	28.3	Science	Comedy	7	Bsc & BA	Bsc & BA	Student	Student	65	45
C#5	39.63	27.0	sport	Religion	4	School Leaver	School Leaver	Student	Student	61	55
C#6	10.0	24.0	Science	Religion	6	Msc & MA	Msc & MA	Student	Student	68	66

Table 1: Two-Step algorithm result table

According to the discussed parameters previously, performed clustering by this algorithm is not really suitable for our purpose, while representative of the characteristics of each cluster has overlapped with other clusters.

5.2. Kohonen Algorithm Result

Deploying this algorithm distributed data in 10 clusters. Distribution form of data in clusters is depicted as below table 2:

Cluster #	Number of Records (%)	Age (avr.)	Favourite Category		Number of likes in This Category (Median)	Education		Job		Marital Status (Single: %)	Sex (Male: %)
			Median	Mode		Median	Mode	Median	Mode		
X=0 Y=0	20.61	23.7	Art & Culture	Comedy & Entertainment	5	School Leaver	PHD	Employee	Employee	75	60
X=0 Y=2	15.81	33.0	Sport	Religion	5	Higher National Diploma	Bsc & BA	Industrialist	Labour	7	65
X=1 Y=0	12.29	29.3	Sacred Defence	Sacred Defence	5	Msc & MA	Bsc & BA	Student	Student	39	65
X=1 Y=2	0.69	26.6	Religion	Religion	4	Msc & MA	Msc & MA	Student	Student	63	59
X=2 Y=0	7.10	27.95	Sacred Defence	Sacred Defence	5	Bsc & BA	Bsc & BA	Student	Student	67	54
X=2 Y=1	5.82	27.24	Religion	Religion	4	High School Diploma	High School Diploma	Student	Student	62	79
X=2 Y=2	10.49	24.1	Science	Religion	6	Msc & MA	Msc & MA	Student	Student	72	65
X=3 Y=0	15.33	26.88	Music	Comedy & Entertainment	7	Bsc & BA	Bsc & BA	Student	Student	55	76
X=3 Y=1	7.66	25.85	News	News	6	High School Diploma	High School Diploma	Student	Student	61	84
X=3 Y=2	4.20	25.9	Science	Religion	4	School Leaver	School Leaver	Student	Student	53	31

Table 2: Kohonen algorithm result table

The same as Two-Step algorithm, this clustering algorithm is not efficient enough for our data set. Clustering does not have normal distribution and of course representative of each cluster has overlapped with other clusters.

5.3. K-means Algorithm Result

In K-mean algorithm which is implemented in Clementine, the number of clusters must be pre-specified. We want to have the most normal distribution in clusters. The output of applying this algorithm for various value of K (number of clusters) are listed in the table 3 as followed:

Title

K	Cluster #	Number of Records (%)	Age (avr.)	Favourite Category		Number of likes in This Category (Median)	Education		Job		Marital Status (Single: %)	Sex (Male: %)
				Median	Mode		Median	Mode	Median	Mode		
K=2	C#1	83.5	26.0	Science & Technology	Religion & Culture	5	School Leaver	Bsc & BA	Student	Master	63	70
	C#2	16.5	33.0	Sport	Religion & Culture	4	Higher National Diploma	Bsc & BA	Industrialist	Labour	38	65
K=3	C#1	55.83	27.7	Science & Technology	Religion & Culture	5	Higher National Diploma	Bsc & BA	Student	Student	52	72
	C#2	16.5	33.0	Sport	Religion & Culture	5	Higher National Diploma	Bsc & BA	Industrialist	Labour	7	65
	C#3	27.67	22.7	Science & Technology	Religion & Culture	6	High School Diploma	Msc & MA	Labour	Employee	81	63
K=4	C#1	43.33	26.9	Sport	Religion	5	School Leaver	School Leaver	Tradesman	Student	59	80
	C#2	11.51	24.2	Science & Technology	Religion	6	Msc & MA	Msc & MA	Student	Student	75	67
	C#3	28.66	25.6	Science & Technology	Religion	5	School Leaver	Bsc & BA	Labour	Labour	61	55
	C#4	16.5	33.0	Sport	Religion	4	Higher National Diploma	Bsc & BA	Labour	Student	45	65
K=5	C#1	24.05	27.3	Art & Culture	Religion	6	Msc & MA	Bsc & BA	Student	Student	53	66
	C#2	11.15	24.2	Science & Technology	Religion	3	Msc & MA	Msc & MA	Student	Student	65	64
	C#3	17.41	26.4	Poetry & Literature	Sacred Defence	6	School Leaver	Bsc & BA	Student	Student	69	66
	C#4	16.5	33	Sport	Religion	4	Theological Seminary	Bsc & BA	Industrialist	Labour	7	65
	C#5	30.89	25.6	Art & Culture	Religion	5	Higher National Diploma	Higher National Diploma	Labour	Employee	15	60
K=6	C#1	16.85	28.4	News	Comedy & Entertainment	5	Msc & MA	Bsc & BA	student	student	36	76
	C#2	10.73	24.1	Religion	Science & Technology	6	Msc & MA	Msc & MA	student	student	79	65

	C#3	13.49	25.9	Food	Sacred Defence	4	School Leaver	Bsc & BA	Employee	Employee	45	48
	C#4	16.5	33	Sport	Religion	4	Higher National Diploma	Bsc & BA	Industrialist	Labour	19	65
	C#5	21.57	25.1	News	Comedy & Entertainment	7	School Leaver	School Leaver	Other	Employee	56	76
	C#6	20.86	26.2	Politics	Religion	5	Higher National Diploma	School Leaver	Tradesman	Student	61	73

Table 3: K-means algorithm result table

As it is seen, users distribution in clusters for K=6 is better than others because with this cluster number, clusters less overlapped and distribution in clusters are almost normal. Also category value and educational level in clusters are distinct. Distributions for k=6, can be seen in following figures 1 to 4 for some different properties.

Cluster overall (Number of records) distribution:

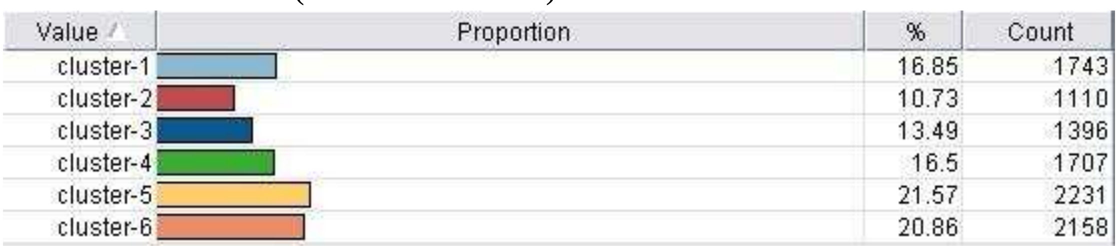


Figure 1: K-means algorithm cluster distribution

Age Distribution:

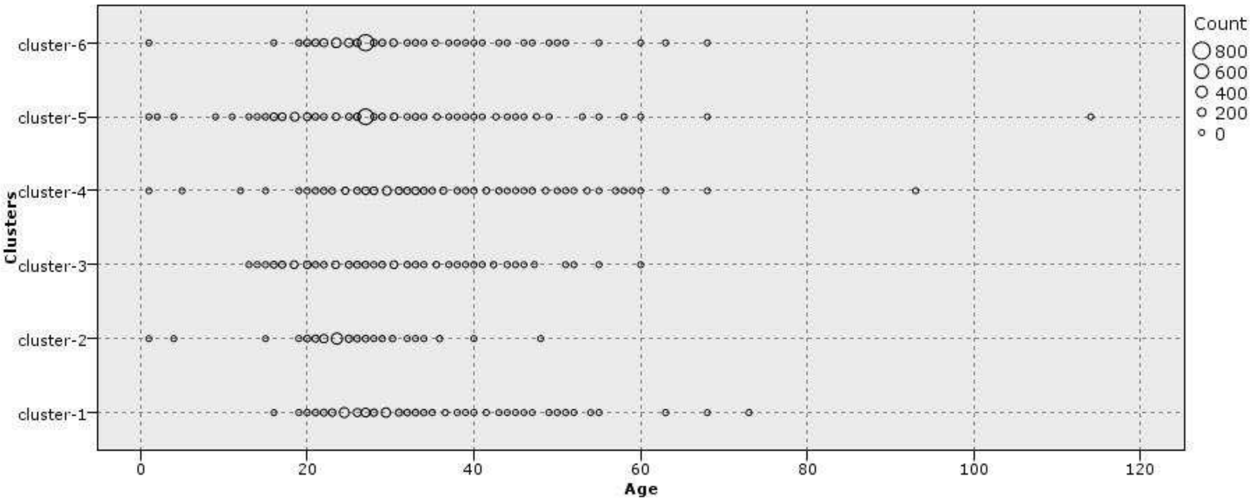


Figure 2: K-means algorithm age distribution

As it is evident based on figure 2, most of users are young but in each category, age range is somehow different from other categories.

Favourite category distribution:

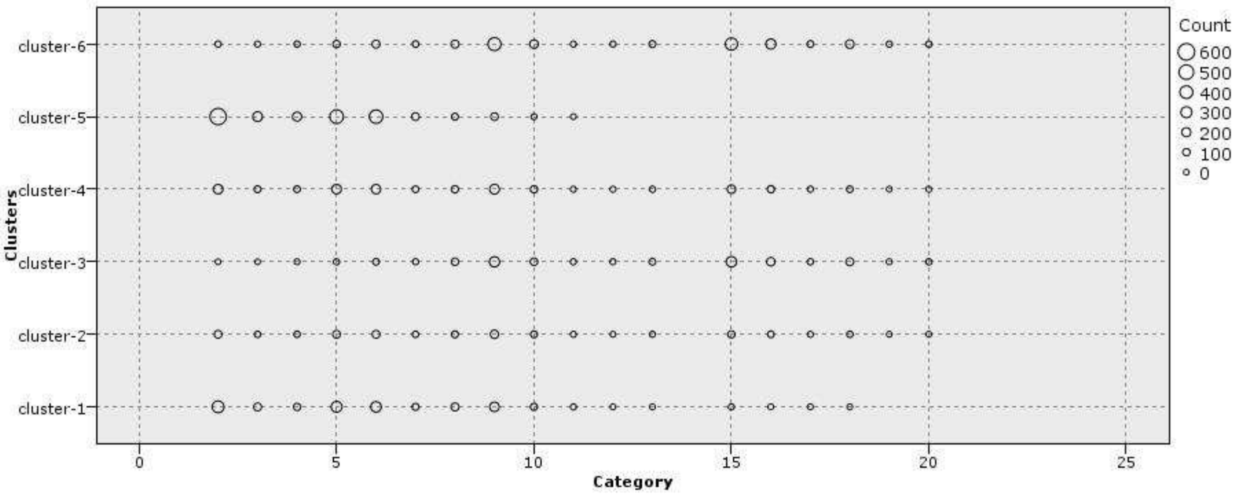


Figure 3: K-means algorithm favourite category distribution

As it is shown in figure 3, favourite category in each cluster may differ greatly in comparison to other clusters. Obviously in all clusters the multitude of users in some categories is far greater than others.

Sex distribution:

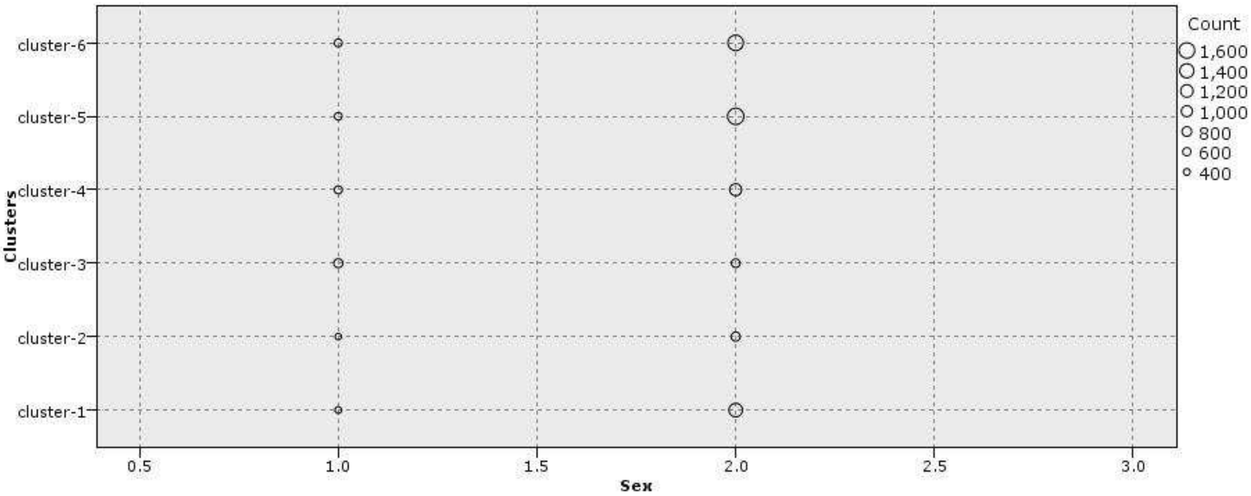


Figure 4: K-means algorithm sex distribution

Based on figure 4 and previous tables, one significant fact is noticeable. In most categories large numbers of users are found to be male whereas in some categories male and female users tend to be the same.

Since we aim to use the clustering results for market segmentation purposes, and want to use it for targeted advertisement, we provide an analytical data for clusters. So we investigate records in each cluster in K=6 in K-mean algorithm. In previous algorithm we cannot find the best normal distribution in the resulted segments. So different algorithm are examined and finally the K-mean algorithm with K=6 is the best one which is described.
In C#1 it seems users are interested in Comedy and Entertainment posts. This cluster are those with higher educational levels. In C#2, users generally follow scientific subjects in this social network. According to their educational level, offering new tools, gadgets and technologies or scientific and religious seminars can be of interest. There are various ranges of ages in C#3, but their common interest is in the Sacred Defence and Food. According to their more free time than others, offering books and magazines or even tours related to mentioned topics can be interesting for them. Users in C#4 are religious. They also have a look at sporting events, so this segment of the market can be considered based on their priorities. Users in C#5 are similar to C#1, this cluster's members have lower educational level, so this algorithm divided them in to tow different clusters. For these

clusters, we can suggest the same propositions as C#1, but it should be noted with less conceptual or complicated content. Users in C#6 are so interested in religion and politics. They are generally students and these users can be considered in religious and cultural fields and promotional content.

6. Discussion and Conclusion

Today's successful businesses must aggressively attack target markets and niche segments, focusing on customer behaviour and their preferences. Online marketers can use the demographic and experience profiles to predict their consumers' segment, future behaviour and interests. Whereas current study increases our understanding about the segments that form the online consumer market according to web use pattern and demographic profiles, it has some limitations. Although scholars highlight problems in the conceptualization and operationalization of segmentation (Dibb and Simkin, 2009; Dolcinar and Lazarevski, 2009; Quinn, 2009), it is a central topic in marketing theory and practice (Wedel and Kamakura, 2000).
Dependent on Tebyan social network's strategy, their need for growth and innovative marketing policies and initiatives, and according to our findings, this study have some meaningful suggestions. If Tebyan wants to expand their users and spread of Tebyan among the target market, they have to invest on advertising and developing marketing policies based on the result of K-means algorithm. While one of the most important missions of Tebyan is religious and cultural development, with respect to the findings, they can focus mainly on young audience who are between 24 to 30 years old, and single men show more preference to this social network in different educational level. While past researches indicated that online shoppers are most likely to be younger males with high internet experience, greater education, and higher income (Li et al., 1999; Sin and Tse, 2002; Swinyard and Smith, 2003; Lyndon Simkin, 2008). One of the influencing factors in previous researches is educational level which is not so important for Tebyan based on our data analysis. But it seems that differences in age and sex are significant amongst Tebyan users; the same as similar researches.
As Srinivasan et al. (2002) examined the antecedents and consequences of customer loyalty in an online context and identified the factors that significantly affect e-loyalty (Lyndon Simkin, 2008). If Tebyan plans to increase their users' loyalty and target to gain more audiences, they can focus on young single men between 20 to 35 years old, which most of them have academic education, students and employees.

As the most of past researches expounded, to develop the users (customers) and to increase previous customers' loyalty, the demographic characteristics are important while this study declares for different and special purpose applying data mining techniques in a special data set of a social network users.

This research, like all, is subject to certain limitations. First, our research and analysis is focussed on special purpose of users´ segmentation in a social network; and the selected case was Tebyan. Further research could target to understand other social networks that exist; Facebook is also routinely used to present consumers with marketing messages which seems are based on target audiences. As the data come from consumers belonging to a large user panel in Iran, researchers can try to validate the findings in other countries and examine the moderating role of culture and country characteristics. To validate these findings in other situations, it is recommended that other researchers do similar investigations on other social networks and their users, especially big and international social networks like Facebook and Twiter, specific use like Instagram or other local social networks like Streetlife in UK and Italylink in Italy.
Also, it can be checked with scholars that how different characteristics can influence the results. Because of mentioned limitation, we have checked some specific characteristics and features which we believe can influence the result, but further specifications can be examined in future studies for different social networks, and with respect to local situations and culture which may influence the results. It is helpful in further studies to make it clear that how these result and changing market strategies affect the outputs for social networks.
We discuss data mining applications from a broader scope, which has some important applications in marketing:

- comprehending customers' needs and predicting their responses to new products/services and communication programs
- identifying customers' loyalty levels and making marketing strategies to retain vulnerable customers
- recognizing unprofitable customers; and
- market segmentation and products/services differentiation (Peacock, 1998; Cheng et al., 2005).

We also claim that these applications can be realized by specific data mining functions and techniques (Campbell et al, 2014).

References

Ahmed, S.R. (2004), "Application of data mining in retail businesses", IEEE International Conference on Information Technology: Coding and Computing, ITCC'04, pp. 44-55.

Title

Ajay Mehra,Stephen P. Borgatti,Scott Soltis,Theresa Floyd,Daniel S. Halgin,Brandon Ofem and,Virginie Lopez-Kidwell (2014), Imaginary Worlds: Using Visual Network Scales to Capture Perceptions of Social Networks, in Daniel J. Brass, Giuseppe (JOE) Labianca, Ajay Mehra, Daniel S. Halgin, Stephen P. Borgatti (ed.) Contemporary Perspectives on Organizational Social Networks (Research in the Sociology of Organizations, Volume 40) Emerald Group Publishing Limited, pp.315 - 336

Alejandra Segura, Christian Vidal-Castro, Víctor Menéndez-Domínguez, Pedro G. Campos, Manuel Prieto, (2011) "Using data mining techniques for exploring learning object repositories", The Electronic Library, Vol. 29 Iss: 2, pp.162 - 180

Ana Kovacevic, VladanDevedzic, Viktor Pocajt, (2010) "Using data mining to improve digital library services", The Electronic Library, Vol. 28 Iss: 6, pp.829 - 843

Anderson, J.C. and Narus, J.A. (2003), "Selectively pursuing more of your *customer's busine*ss", Sloan Management Review, Vol. 44 No. 3, pp. 42-9.

Berson, A., Smith, S. and Thearling, K. (1999), Building Data Mining Applications for CRM, McGraw-Hill Professional, New York, NY.

Boyd, D.M. and Ellison, N.B. (2007), "Social network sites: definition, history, and scholarship", Journal of Computer-Mediated Communication, Vol. 13 No. 1, pp. 210-230.

Chen, L. and Sakaguchi, T. (2000), "Data mining methods, applications, and tools", Information Systems Management, Vol. 17 No. 1, p. 65.

Cheng, B., Chang, C. and Liu, I. (2005), "Enhancing care services quality of nursing homes using data mining", Total Quality Management & Business Excellence, Vol. 16 No. 5, pp. 575-96.

Cheng, J., Wong, S.H.Y., Yang, H. and Lu, S. (2007), "SmartSiren: virus detection and alert for smartphones", paper presented at MobiSys'07, San Juan, Puerto RicoMobiSys'07, June 11-14, San Juan, Puerto Rico.

Chopoorian, J.A., Witherell, R., Khalil, O.E.M. and Ahmed, M. (2001), "Mind your business by mining your data", Advanced Management Journal, Vol. 66 No. 2, p. 45.

Colin Campbell,Carla Ferraro,Sean Sands, (2014) "Segmenting consumer reactions to social network marketing", European Journal of Marketing, Vol. 48 Iss: 3/4, pp.432 - 452

comScore (2012a), It's a Social World: Top 10 Need-to-Knows About Social Networking and Where It's Headed, available at: www.comscore.com/Press_Events/Presentations_Whitepapers/2011/it_is _a_social _world_top_10_need-to-knows_about_social_networking (accessed 22 July 2012).

comScore (2012b), Introducing Mobile Metrix 2 Insight into Mobile Behavior, available at: www.

comscore.com/Press_Events/Press_Releases/2012/5/Introducing_Mobile_Metrix_2_Insight_into_Mobile_Behavior/ (accessed 22 July 2012).
Craft, S.H. (2004), "The international consumer market segmentation managerial decision-making process", SAM Advanced Management Journal, Vol. 69 No. 3, pp. 40-8.
Desislava G. Budeva, Michael R. Mullen, (2014) "International market segmentation: Economics, national culture and time", European Journal of Marketing, Vol. 48 Iss: 7/8, pp.1209 - 1238
Devedzic, V. (2001), "Knowledge discovery and data mining in databases", in Chang, S.K. (Ed.), Handbook of Software Engineering and Knowledge Engineering Vol. 1 – Fundamentals, World Scientific Publishing, Singapore, pp. 615-37.
Dibb, S. and Simkin, L. (2008), Market Segmentation Success: Making It Happen!, The Haworth Press, New York, NY.
Dibb, S. and Simkin, L. (2009), "Implementation rules to bridge the theory/practice divide in market segmentation", Journal of Marketing Management, Vol. 25 Nos 3/4, pp. 375-96.
Dibb, S., Stern, P. and Wensley, R. (2002), "Academic views on market segmentation", Marketing Intelligence & Planning, Vol. 20, April/March, pp. 113-19.
Dolcinar, S. and Lazarevski, K. (2009), "Methodological reasons for the theory/practice divide in marketing segmentation", Journal of Marketing Management, Vol. 25 Nos 3/4, pp. 357-73.
Eastman, J. and Iyer, R. (2004), "The elder*ly's uses and attitudes towards the* internet", Journal of Consumer Marketing, Vol. 21 Nos 2/3, pp. 208-20.
Fayyad, U.M., Piatsky Shapiro, G. and Smyth, P. (1996), "From data mining to knowledge discovery in data base", AI Magazine, Fall, pp. 37-54.
Foster, M., West, B. and Francescucci, A. (2011), "Exploring social media user segmentation and online brand profiles", Journal of Brand Management, Vol. 19 No. 1, pp. 4-17.
Fournier, S. and Avery, J. (2011), "The uninvited brand", Business Horizons, Vol. 54 No. 3, pp. 193-207.
Gefen, D. and Straub, D. (1997), "Gender differences in perception and adoption of e-mail: an extension to the technology acceptance model", Management Information Systems Quarterly, Vol. 21 No. 4, pp. 389-400.
Han, J. and Kamber, M. (2006), Data Mining Concepts and Techniques, Morgan Kaufmann, San Francisco, CA, pp. 4-27.
Hanna, R., Rohm, A. and Crittenden, V.L. (2011), "*We're all connected: the power* of the social media ecosystem", Business Horizons, Vol. 54 No. 3, pp. 265-273.
Hassan, S.S. and Craft, S.H. (2005), "Linking global market segmentation decisions with strategic positioning options", Journal of Consumer Marketing, Vol. 22 No. 2, pp. 81-9.
Hunt, S. (1991), Modern Marketing Theory, South-Western Publishing, Cincinnati, OH.

Ip, R.K.F. and Wagner, W. (2008), “Weblogging: a study of social computing and its impact on organizations”, Decision Support Systems, Vol. 45, pp. 242-250.

JayanthiRanjan, (2009) "Data mining in pharma sector: benefits", International Journal of Health Care QualityAssurance, Vol. 22 Iss: 1, pp.82 – 92

JayanthiRanjan, Vishal Bhatnagar, (2009) "A holistic framework for mCRM – data mining perspective", Information Management & Computer Security, Vol. 17 Iss: 2, pp.151 – 165

Kantardzic, M. (2003), Data Mining: Concepts, Models, Methods, and Algorithms, John Wiley & Sons, Hoboken, NJ.

Kaplan, A.M. and Haenlein, M. (2010), “Users of the world, unite! The challenges and opportunities of social media”, Business Horizons, Vol. 53 No. 1, pp. 59-68.

Kaplan, A.M. and Haenlein, M. (2011a), “Two hearts in three-quarter time: how to waltz the social media/viral marketing dance”, Business Horizons, Vol. 54 No. 3, pp. 253-263.

Kau, A.K., Tang, Y.E. and Ghose, S. (2003), “Typology of online shoppers”, Journal of Consumer Marketing, Vol. 20 No. 2, pp. 139-56.

Kietzmann, J., Hermkens, K., McCarthy, I. and Silvestre, B. (2011), “Social media? Get serious! Understanding the functional building blocks of social media”, Business Horizons, Vol. 54 No. 3, pp. 241-251.

Konus, U., Verhoef, P.C. and Neslin, S.A. (2008), “Multichannel shopper segments and their covariates”, Journal of Retailing, Vol. 84 No. 4, pp. 398-413.

Kushwaha, T.L. and Shankar, V. (2008), Single channel vs. multichannel retail customers: correlates and consequences, working paper, Texas A&M University, College Station, TX.

Larose, D. (2005), Discovering Knowledge in Data: An Introduction to Data Mining, John Wiley & Sons, Hoboken, NJ.

Lejeune, M. (2001), “Measuring the impact of data mining on churn management”, Internet Research, Vol. 11 No. 5, pp. 295-304.

Li, C. and Bernoff, J. (2008), Groundswell: Winning in a World Transformed by Social Technologies, Harvard Business Press, Boston, MA.

Li, H., Kuo, C. and Russell, M.G. (1999), “The impact of perceived channel *utilities, shopping orientations, and demographics on the consumer’s* online buying behavior”, Journal of Computer Mediated Communication, Vol. 5 No. 2, p. 36.

Lyndon Simkin, (2008) "Achieving market segmentation from B2B sectorisation", Journal of Business & Industrial Marketing, Vol. 23 Iss: 7, pp.464 - 474

Mangold, W. and Faulds, D. (2009), “Social media: the new hybrid element of the promotion mix”, Business Horizons, Vol. 52 No. 4, pp. 357-365.

Mathwick, C. (2001), “Understanding the online consumer: a typology of online relational norms and behavior”, Journal of Interactive Marketing, Vol. 16 No. 1, pp. 40-55.

Mostafa, M.M. (2006), “An empirical investigation of Egyptian consumers usage patterns and perceptions of the internet”, International Journal of Management, Vol. 23 No. 2, pp. 243-61.

Muhammad Aljukhadar, Sylvain Senecal, (2011) "Segmenting the online consumer market", Marketing Intelligence & Planning, Vol. 29 Iss: 4, pp.421 - 435

Nielsen (2011), State of the Media: The Social Media Report, Nielsen Consumer Research, available at: http://blog.nielsen.com/nielsenwire/social/ (accessed 18 October 2011).

Palmer, R.A. and Millier, P. (2004), “Segmentation: identification, intuition and implementation”, Industrial Marketing Management, Vol. 33 No. 8, pp. 779-85.

Paola Carone,Simona Panaro, (2014) "New images of city through the social network", International Journal of Web Information Systems, Vol. 10 Iss: 2, pp.209 - 223

Peacock, P.R. (1998), “Data mining in marketing: part 1”, Marketing Management, Vol. 6 No. 4, pp. 8-18.

Peter Groenewegen and,Christine Moser (2014), Online Communities: Challenges and Opportunities for Social Network Research, in Daniel J. Brass, Giuseppe (JOE) Labianca, Ajay Mehra, Daniel S. Halgin, Stephen P. Borgatti (ed.) Contemporary Perspectives on Organizational Social Networks (Research in the Sociology of Organizations, Volume 40) Emerald Group Publishing Limited, pp.463 - 477

Pew Internet and American Life Project (2010), available at: www.pewinternet.org/Static-Pages/Trend-Data/Online-Activites-Total.aspx (accessed January 8, 2011).

Quinn, L. (2009), “Market segmentation in managerial practice: a qualitative examination”, Journal of Marketing Management, Vol. 25 Nos 3/4, pp. 253-72.

Ranjan, J. and Bhatnagar, V. (2008b), “Data mining tools-a CRM perspectives”, International Journal of Electronic CRM, Vol. 2 No. 4, pp. 315-31.

Rheingold, H. (2002), Smart Mobs: The Next Social Revolution, Basic Books, New York, NY.

Riegner, C. (2007), “Word of mouth on the web: the impact of Web 2.0 on consumer purchase decisions”, Journal of Advertising Research, Vol. 47 No. 4, pp. 436-437.

Sandra S. Liu, Jie Chen, (2009) "Using data mining to segment healthcare markets from patients' preferenceperspectives", International Journal of Health Care Quality Assurance, Vol. 22 Iss: 2, pp.117 - 134

Sang Jun Lee, KengSiau, (2001) "A review of data mining techniques", Industrial Management & DataSystems, Vol. 101 Iss: 1, pp.41 - 46

SAS Institute Inc. (2002), Applying Data Mining Techniques Using Enterprise Miner, SAS Institute Inc., Cary, NC.

Title

Sin, L. and Tse, A. (2002), "Profiling internet shoppers in Hong Kong: demographic, psychographic, attitudinal and experiential factors", Journal of International Consumer Marketing, Vol. 15 No. 1, pp. 7-30.

Srinivasan, S., Anderson, R. and Ponnavolu, K. (2002), "Consumer loyalty in e-commerce: an exploration of its antecedents and consequences", Journal of Retailing, Vol. 78 No. 1, pp. 41-50.

Stephen P. Borgatti,Daniel J. Brass and,Daniel S. Halgin(2014), Social Network Research: Confusions, Criticisms, and Controversies, in Daniel J. Brass, Giuseppe (JOE) Labianca, Ajay Mehra, Daniel S. Halgin, Stephen P. Borgatti (ed.) Contemporary Perspectives on Organizational Social Networks (Research in the Sociology of Organizations, Volume 40) Emerald Group Publishing Limited, pp.1 - 29

Strebel, J., Erdem, T. and Swait, J. (2004), "Consumer search in high technology markets: exploring the use of traditional information channels", Journal of Consumer Psychology, Vol. 14, pp. 96-104.

Swinyard, W.R. (1996), "The hard core and Zen riders of Harley Davidson: a market-driven segmentation analysis", Journal of Targeting, Measurement and Analysis for Marketing, Vol. 4 No. 4, pp. 337-62.

Swinyard, W.R. and Smith, S.M. (2003), "Why *people (don't) shop online: a* lifestyle study of the internet consumer", Psychology & Marketing, Vol. 20 No. 7, pp. 567-97.

Thelen, S., Mottner, S. and Berman, B. (2004), "Data mining: on the trail to marketing gold", Business Horizons, Vol. 47 No. 6, pp. 25-32.

Vandermerwe, S. (2004), "Achieving deep customer focus", Sloan Management Review, Vol. 45 No. 3, pp. 25-34.

Wasko, M.M. and Faraj, S. (2000a), "It is what one does: why people participate and help others in electronic communities of practice", Journal of Strategic Information Systems, Vol. 9, pp. 155-173.

Wasko, M.M. and Faraj, S. (2000b), "Why should I share? Examining social capital and knowledge contribution in electronic networks of practice", MIS Quarterly, Vol. 29 No. 1, pp. 35-57.

Wedel, M. and Kamakura, W. (1999), Market Segmentation: Conceptual and Methodological Foundations, Kluwer Academic Publishers, Boston, MA.

Wedel, M. and Kamakura, W. (2000), Market Segmentation: Conceptual and Methodological Foundations, Kluwer Academic Publishers, Boston, MA.

Wedel, M. and Kamakura, W. (2002), "Introduction to a special issue in marketing segmentation", International Journal of Research in Marketing, Vol. 18 No. 3, pp. 181-183.

Wedel, M. and Steenkamp, J.-B.E.M. (1989), "Fuzzy clusterwise regression approach to benefit segmentation", International Journal of Research in Marketing, Vol. 6, pp. 241-58.

Weinberg, B.D. and Pehlivan, E. (2011), "Social spending: managing the social media mix", Business Horizons, Vol. 54 No. 3, pp. 275-282.

Weinstein, A. (2004), Handbook of Market Segmentation, The Haworth Press, New York, NY.

Author

Wiertz, C. and DeRuyter, K. (2007), “Beyond the call of duty: why consumers contribute to firm-hosted commercial online communities”, Organization Studies, Vol. 28 No. 3, pp. 347-376.

Cluster #	Number of Records (%)	Age (avr.)	Favourite Category		Number of likes in this category (Median)	Education		Job		Marital Status (Single: %)	Sex (Male: %)
			Median	Mode		Median	Mode	Median	Mode		
C#1	15.66	32.0	Science	Religion	4	Higher National Diploma	BA & Bsc	Industrialist	Labour	19	66
C#2	4.61	40.8	Culture	Comedy	6	Higher National Diploma	Bsc & BA	Student	Student	7	60
C#3	15.16	19.5	Sport	Comedy	5	High School Diploma	High School Diploma	Student	Student	86	62
C#4	14.93	28.3	Science	Comedy	7	Bsc & BA	Bsc & BA	Student	Student	65	45
C#5	39.63	27.0	sport	Religion	4	School Leaver	School Leaver	Student	Student	61	55
C#6	10.0	24.0	Science	Religion	6	Msc & MA	Msc & MA	Student	Student	68	66

Table 4~~1~~: Two-Step algorithm result table

Cluster #	Number of Records (%)	Age (avr.)	Favourite Category		Number of likes in This Category (Median)	Education		Job		Marital Status (Single: %)	Sex (Male: %)
			Median	Mode		Median	Mode	Median	Mode		
X=0 Y=0	20.61	23.7	Art & Culture	Comedy & Entertainment	5	School Leaver	PHD	Employee	Employee	75	60
X=0 Y=2	15.81	33.0	Sport	Religion	5	Higher National Diploma	Bsc & BA	Industrialist	Labour	7	65
X=1 Y=0	12.29	29.3	Sacred Defence	Sacred Defence	5	Msc & MA	Bsc & BA	Student	Student	39	65
X=1 Y=2	0.69	26.6	Religion	Religion	4	Msc & MA	Msc & MA	Student	Student	63	59
X=2 Y=0	7.10	27.95	Sacred Defence	Sacred Defence	5	Bsc & BA	Bsc & BA	Student	Student	67	54
X=2 Y=1	5.82	27.24	Religion	Religion	4	High School Diploma	High School Diploma	Student	Student	62	79
X=2 Y=2	10.49	24.1	Science	Religion	6	Msc & MA	Msc & MA	Student	Student	72	65
X=3 Y=0	15.33	26.88	Music	Comedy & Entertainment	7	Bsc & BA	Bsc & BA	Student	Student	55	76
X=3 Y=1	7.66	25.85	News	News	6	High School Diploma	High School Diploma	Student	Student	61	84
X=3 Y=2	4.20	25.9	Science	Religion	4	School Leaver	School Leaver	Student	Student	53	31

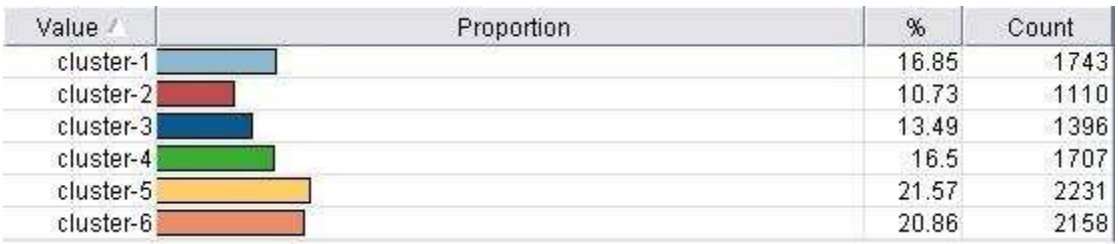
Table 5~~2~~: Kohonen algorithm result table

K	Cluster #	Number of Records (%)	Age (avr.)	Favourite Category		Number of likes in This Category (Median)	Education		Job		Marital Status (Single: %)	Sex (Male: %)
				Median	Mode		Median	Mode	Median	Mode		
K=2	C#1	83.5	26.0	Science & Technology	Religion & Culture	5	School Leaver	Bsc & BA	Student	Master	63	70
	C#2	16.5	33.0	Sport	Religion & Culture	4	Higher National Diploma	Bsc & BA	Industrialist	Labour	38	65
K=3	C#1	55.83	27.7	Science & Technology	Religion & Culture	5	Higher National Diploma	Bsc & BA	Student	Student	52	72
	C#2	16.5	33.0	Sport	Religion & Culture	5	Higher National Diploma	Bsc & BA	Industrialist	Labour	7	65
	C#3	27.67	22.7	Science & Technology	Religion & Culture	6	High School Diploma	Msc & MA	Labour	Employee	81	63
K=4	C#1	43.33	26.9	Sport	Religion	5	School Leaver	School Leaver	Tradesman	Student	59	80
	C#2	11.51	24.2	Science & Technology	Religion	6	Msc & MA	Msc & MA	Student	Student	75	67
	C#3	28.66	25.6	Science & Technology	Religion	5	School Leaver	Bsc & BA	Labour	Labour	61	55
	C#4	16.5	33.0	Sport	Religion	4	Higher National Diploma	Bsc & BA	Labour	Student	45	65
K=5	C#1	24.05	27.3	Art & Culture	Religion	6	Msc & MA	Bsc & BA	Student	Student	53	66
	C#2	11.15	24.2	Science & Technology	Religion	3	Msc & MA	Msc & MA	Student	Student	65	64
	C#3	17.41	26.4	Poetry & Literature	Sacred Defence	6	School Leaver	Bsc & BA	Student	Student	69	66
	C#4	16.5	33	Sport	Religion	4	Theological Seminary	Bsc & BA	Industrialist	Labour	7	65
	C#5	30.89	25.6	Art & Culture	Religion	5	Higher National Diploma	Higher National Diploma	Labour	Employee	15	60
K=6	C#1	16.85	28.4	News	Comedy & Entertainment	5	Msc & MA	Bsc & BA	student	student	36	76
	C#2	10.73	24.1	Religion	Science & Technology	6	Msc & MA	Msc & MA	student	student	79	65
	C#3	13.49	25.9	Food	Sacred Defence	4	School Leaver	Bsc & BA	Employee	Employee	45	48

Title

	C#4	16.5	33	Sport	Religion	4	Higher National Diploma	Bsc & BA	Industrialist	Labour	19	65
	C#5	21.57	25.1	News	Comedy & Entertainment	7	School Leaver	School Leaver	Other	Employee	56	76
	C#6	20.86	26.2	Politics	Religion	5	Higher National Diploma	School Leaver	Tradesman	Student	61	73

Table 6~~3~~: K-means algorithm result table



Value	Proportion	%	Count
cluster-1		16.85	1743
cluster-2		10.73	1110
cluster-3		13.49	1396
cluster-4		16.5	1707
cluster-5		21.57	2231
cluster-6		20.86	2158

Figure 5~~1~~: K-means algorithm cluster distribution

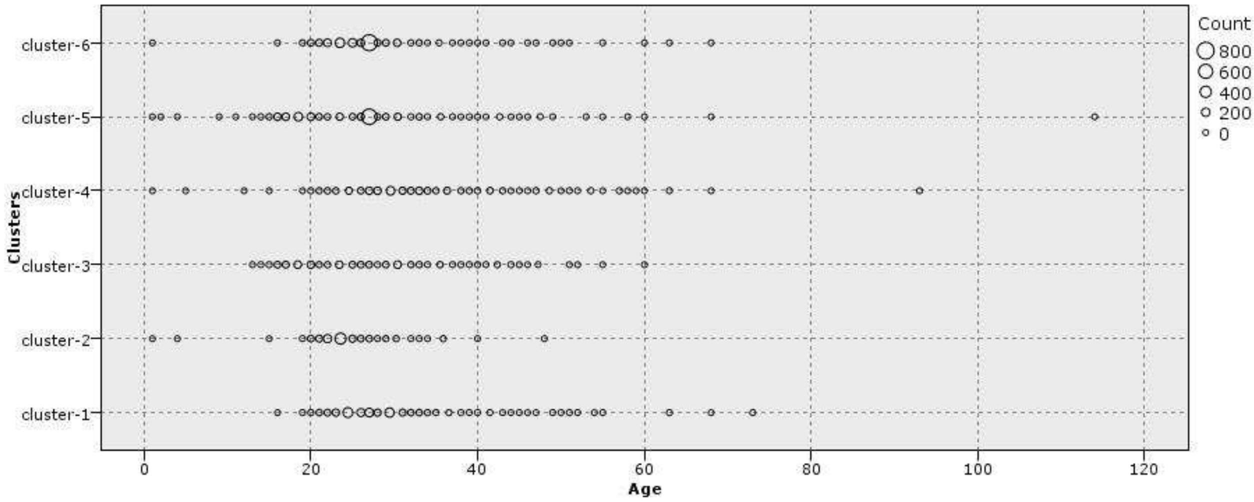


Figure 6~~2~~: K-means algorithm age distribution

Author

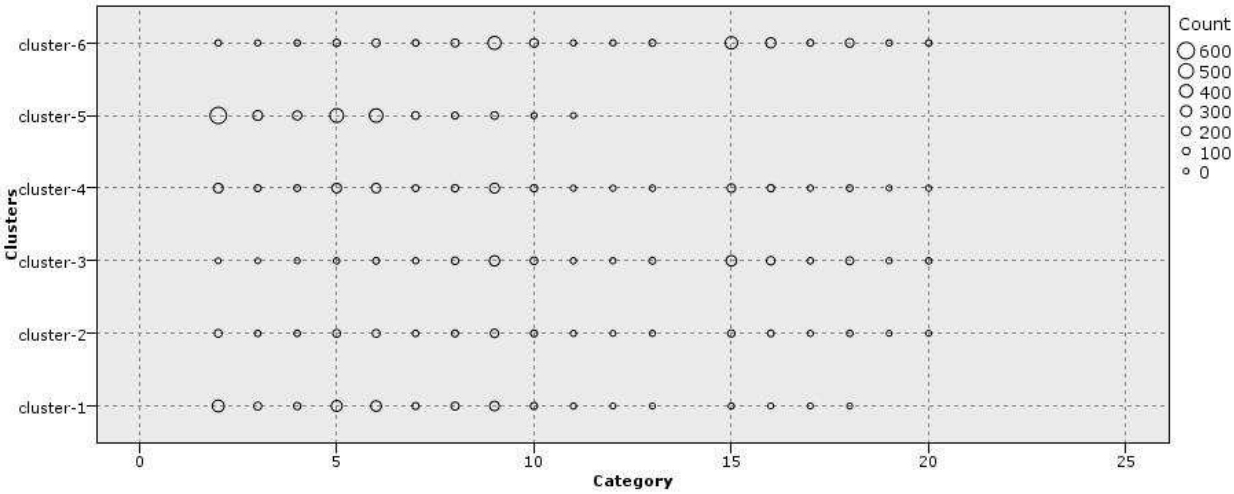


Figure 7~~3~~: K-means algorithm favourite category distribution

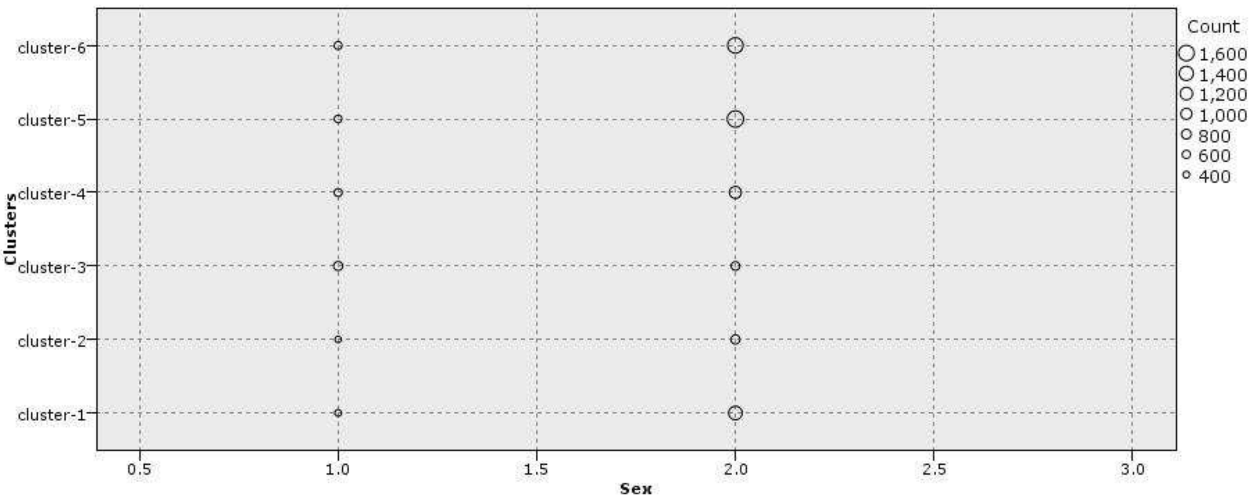


Figure 8~~4~~: K-means algorithm sex distribution