

Online Shopping Adoption Factors in Kuwait Market based on Data Mining Rough Set Approach

Luai A. Al-Shalabi
Arab Open University
Faculty of Computer Studies
Kuwait

ABSTRACT

This study applied data mining process to provide an insight about factors which led to the adoption of online shopping in medium and Large-sized shopping centers (MLShCs) in Kuwait. This research is focused on proposing the high quality environmental factors which affect the success of online shopping among (MLShCs) in Kuwait and ignoring the low impact factors that may cost a lot of money. It also researches the behavior of the customers and their desire to buy directly from the physical store, the online shopping website, or both. This was achieved by distributing a questionnaire, collecting the data in dataset, cleaning the data, minimizing the dimension of the dataset vertically by applying the rough set feature selection technique, and building a classification model. The result of the previous process is the key to examine the success of online shopping in MLShCs. This study could work as the decision maker for those new investors in Kuwait who are thinking of establishing new shopping business and advise them to go for physical store, online shopping website, or both. It also advises current online shopping business to improve their infrastructure, website's design and availability, and security issues. The proposed approach performs effectively and generates interested results.

Keywords

E-commerce, Data Mining, feature selection, classifiers.

1. INTRODUCTION

Electronic commerce which known as e-commerce can be defined as the use of the internet to conduct business transactions nationally or internationally (WTO, 1998). E-commerce has come to take on two important roles; first as a more effective and efficient conduit and aggregator of information, and second, as a potential mechanism for the replacement of many economic activities once performed within a business enterprise by those that can be done by outside suppliers that compete with each other to execute these activities (ECLAC, 2002).

With the rapid development of science and technology in today's society, the country's development is inseparable with the development of IT industry (Yu Zhang *et al.* 2013), and e-commerce industry is the emerging industry in the IT industry (Ramakrishnan Ramanathan *et al.* 2012).

E-commerce offers unprecedented opportunities to both developing and developed countries. In the short run, the gains are likely to be concentrated in developed countries but, in the long run, developing countries have more to benefit. In the short run, developing countries lack the infrastructure necessary to take full advantage of Internet. But in the long run, they can leap frog, skipping some of the stages in the development of information technology through which developed countries have had to pass (Panagariya, A. *et al.*

2000). One impact for e-commerce is to intensify competition and producing benefits to consumers in lower prices and more choices (Malkawi, B. H., 2007).

To clearly analyze the degree of e-commerce's impact on economy it is important to find out the advantages and problems in e-commerce development, adjust the e-commerce industry structure, make greater and more positive contribution to the national economy development (Nuray Terzi, 2011; Dave Chaffey, 2011). Online shopping is a form of e-commerce which allows consumers to directly buy goods or services from a seller over the internet using a web browser.

Customers can walk-in in order to buy goods or they may use the web to do so. The first option may cost effort, road risk, road frustration, waste of time, and so on beside the goods cost while the second option may save all the mentioned costs. Web-based selling is mainly focused on the following two categories: companies which have their own store and at same time they put their catalogues online to allow customers to buy using the web and companies which have started from scratch using the web as their only sales medium and they do not have any physical shopping store opened for customers.

The main contributions of this article are:

- A literature review about e-commerce's impact on economy is presented.
- A literature review about feature reduction methods is presented.
- Identified the main features that influence the online shopping service quality positively and ignored other features that hinder the success of it.
- Build a classification data mining model for medium and large markets in Kuwait which could help in improving the online shopping. Three different models were used and compared based on accuracy and coverage.
- Deep analysis of the main features identified and their influence on customers and markets in Kuwait was conducted. This is done by studying the behavior of customers and note how they perceive online shopping.

Recommendations were introduced to improve online shopping in Kuwait based on the detailed study and the data mining classification model built.

2. E-COMMERCE: AN OVERVIEW

All material on each page should fit within a rectangle of 18 x Information technology revolution is the key of the creation of e-commerce. It is considered an important part of e-commerce and it is widely used in the word trading business.

The internet and e-commerce are transforming the way firms operate by redefining how back-end operations – product design and development, procurement, production, inventory, distribution, aftersales service support, and even marketing are conducted. In this process, the Internet and e-commerce alter the roles and relationships of various parties, fostering new supply networks, services and business models. The end results are efficiency improvements, better asset utilization, faster time to market, reduction in total order fulfillment times, and enhanced customer service (ECLAC, 2002).

Studies suggest a number of factors which lead to more positive evaluations of e-commerce sites, including the characteristics of the organization running the site (Jarvenpaa, S.L., 2000), and the interface characteristics of the sites themselves (Kim, J. and Moon, J.Y., 1998; Egger, F.N., 2000)

Around eighty percent of those transactions were between one business and another. According to Inc. Magazine although the US and Canada lead the world in ecommerce spending, other countries are increasingly shopping online. By 2014, global ecommerce spending is projected to increase more than 90 percent. A sizable portion of that growth is expected to come from Latin America, where the amount spent online is projected to more than double. If these projections are accurate, annual ecommerce spending, in billions for 2014 will be North America \$202.8; Western Europe \$166.5; Asia-Pacific \$93.2; Latin America \$27.1; Eastern Europe & Russia \$27.0; Australia \$4.9; Africa & the Middle East \$3.0 (Inc. Magazine - December 2010/January 2011).

In previous research, different models have been developed to measure customer perception of service quality such as SERVQUAL. SERVQUAL model was developed by Parasuraman et al. in 1988 (Parasuraman, A. *et al.* 1988). It has been widely tested as a means of measuring customer perceptions of service quality. It consists of five dimensions, namely tangibles, reliability, responsiveness, assurance and empathy. During the past decade, SERVQUAL model has been tested for measuring service quality in e-commerce settings (Kuo, Y.F., 2003; Negash, S. *et al.* 2003). Rough set theory is an important mathematical theory that is used to find the most important features (called reduct) which represents informational interesting. This reduct is important to increase the quality of an application or service under study. If the application or the service is of high quality then the accuracy of that application or service will be high. The features of rough set theory will be used in this research article in order to build a data mining classification model for medium and large markets in Kuwait. This model will be used to test if Kuwait market is ready to use online shopping successfully or not. If yes, then the given features of the model are to be taken into account in term of focusing on, developing, or others and this will lead to the improvement and the success of e-commerce and online shopping in Kuwait. Otherwise, recommendations should be given to handle the obstacles.

Al-Hudhaif and others highlighted the factors contributing to the adoption of ecommerce in Kingdom of Saudi Arabia (Al-Hudhaif, S. and Alkubeyyer, A., 2011). Oreku *et al.* help in identifying the important factors in implementation of electronic commerce in East Africa (Oreku, G.S. *et al.* 2011). The study suggests that factors such as culture, economic condition, consumer behavior, purchasing power parity have direct bearing on the implementation of ecommerce. Radovilsky and Hegde (2004) and Kevin Zhu (2004) emphasized in their studies on the technological competence of the firm in the diffusion of e-commerce.

A growing number of Kuwaiti businesses are expressing interest in e-commerce. A study done by Arab Advisors Group (2008) indicated that the percentage of spread of e-commerce in Kuwait reached 10.7% in 2006. The same study concluded that Kuwait's e-commerce users reported a preference for payments through credit cards with 72.3% of e-commerce users reporting credit cards as their e-commerce method of payment.

Al Fadhli has tried to identify the factors that affect electronic commerce applications in Kuwait (Al-Fadhli, S., 2011). He has tried to find out how consumers in Kuwait perceive e-commerce sites, or their subsequent online shopping behavior. It also has tried to identify the main influence on online shopping service quality.

This comprehensive study was made to improving e-commerce in Kuwait shopping sector and then to help new investors to decide using web-based selling or not for their new shopping business. This study was based on different models of data collection, data cleaning, data coding, and data processing.

3. METHODOLOGY

The main objective is to identify the major factors which lead new investors to go for store or online shopping business. The factors also help current companies to improve their business and their success by focusing on the important factors and ignore other factors that impede the success of the business and also to examine the statistical significance of these factors. For this, rough set reduction technique which is one of the data mining techniques that identify the main significant factors was used. The questionnaire was designed on the basis of an extensive literature review and interviews with many IT managers in different companies

3.1 Rough Set Theory

Rough set theory is a mathematical tool to deal with uncertainty (Pawlak Z., and Skowron A., 2007). It can provide a tool for discovering relationships between records and decisions. So, the dataset can be reduced to get the minimum representation in terms of decision. Data mining was used in medicine (Al-Shalabi L., 2009), Learning (Al-Shalabi L., 2016), crime (Al-Shalabi L., 2017), and so on. The main basic concepts of rough set theory as explained by Pawlak are introduced.

Approximation of sets: The lower approximation of X is the collection of objects that can be classified with full certainty as members of the set X using the attribute set A. The upper approximation of X is the collection of objects that possibly are classified as members of the set X. The boundary region includes the objects that cannot be classified with certainty to be inside or outside X, using the same attribute set A.

Let $P \subseteq A$ and $Y \subseteq U$. The P -lower approximation of Y, denoted by $\underline{P}Y$, and P -upper approximation of Y, denoted by $\overline{P}Y$, are defined as:

$$PY = \cup\{X \in U | P: X \subseteq Y\} \quad (1)$$

$$\overline{P}Y = \cup\{X \in U | P: X \cap Y = \emptyset\} \quad (2)$$

The P -boundary of set Y is defined as:

$$Bnp(Y) = \overline{P}Y - \underline{P}Y \quad (3)$$

Set $\underline{P}Y$ is the set of all objects from U which can be certainly classified as elements of Y, employing the set of attributes P. Set $\overline{P}Y$ is the set of objects from U which can be possibly classified as elements of Y, using the set of attributes P. The

set $Bnp(Y)$ is the set of objects which cannot be certainly classified as element of Y using the set of attributes P only.

Rough classification: Let S be an information system, $P \subseteq A$, and let $Y = \{Y_1, Y_2, \dots, Y_n\}$ be a partition of U . By the P -lower approximation of Y in S it means the sets $\underline{P}Y = (\underline{P}Y_1, \underline{P}Y_2, \dots, \underline{P}Y_n)$.

The coefficient:

$$\gamma P(Y) = \sum_{i=1}^n \text{card}(\underline{P}Y_i) / \text{card}(U) \quad (4)$$

is called the quality of classification. It expresses the ratio of all P -correctly classified objects to all objects in the system.

Reduction and dependency of attributes: In the information system S , the minimal subset $R \subseteq P \subseteq A$ such that $\gamma R(Y) = \gamma P(Y)$ is called Y -reduct of P and denoted by $REDY(P)$. Notice that an information system may have more than one Y -reduct. Intersection of all Y -reducts is called the Y -core of P denoted by $COREY(P)$, i.e. $COREY(P) = \bigcap REDY(P)$.

Discovering dependencies among attributes is the primary importance in the rough set approach to knowledge analysis. Set of attributes $Q \subseteq A$ depends on set of attributes $P \subseteq A$, denoted by $P \rightarrow Q$, if each equivalence class of the equivalence relation generated by P is included in some equivalence class generated by Q , i.e.

$$P \rightarrow Q \text{ iff } IND(P) \subseteq IND(Q) \quad (5)$$

3.2 Data collection

A questionnaire and many interviews were undertaken for collection of data. This research utilized online questionnaire to investigate factors affect online shopping in Kuwait. The online questionnaire was available to selected respondents. Designing the questionnaire items in preparation for data collection was the first stage of data collection. A total of 1032 customers were participating in this study.

The questionnaire consists of three parts. Part 1 consists of questions about respondents' personal information (such as gender, age, education level, occupation, income, etc). Part 2 consists of questions about respondents' internet environment and online purchasing experience as well as the shopping website itself (such as how often do they have online shopping, how much do they usually spend on the online shopping per month, what type of products do they typically buy, have they had a dissatisfying experience with a web site, does the design of the website affect the online buying decision, etc.). Part 3 consists of questions about respondents' privacy and security issues (such as How comfortable they buy products online, How confident they are when they submit their personal information on the online shopping system, How often do privacy concerns prevent them from buying online).

Interviews were carried out with owners and managers of some important online shopping business in Kuwait. The purpose of these interviews is to determine the obstacles facing their business. A set of questions were developed to ensure coverage of all issues related to online shopping growth.

3.3 Data preprocessing

Data pre-processing is an important step in the data mining process which includes data cleaning, missing values, normalization, transformation, feature extraction, etc.

Analyzing data that has not been carefully screened for such problems can produce misleading results. Thus, the representation and quality of data Σ is first and foremost before running an analysis (Pyle, D., 1999). In this work, missing data, data coding, and feature selection (reduction of the data set) will be highlighted and resolved.

Missing data might occur because the value is not relevant to a particular case, could not be recorded when the data was collected, or is ignored by users because of privacy concerns (Al-Shalabi, L. *et al.* 2006). The problem of missing values and solutions for this problem has been investigated long time ago (Quinlan J.R., 1989). Dempster *et al.* (1977) proposed methods for solving missing data in datasets. They summarized these methods in three categories. First: Ignoring and discarding data by deleting all records that have missing data. Second: Parameter estimation which is used to find the parameters for the complete data. Third: Imputation technique which replaces the missing values based on estimated values that are the most probable values. The First method will be used in this study.

To identify the important features that affect the online shopping in Kuwait, a two-phased process has been used. In the first phase, data cleaning and data coding models were adopted. The process of deleting all rows with missing values was taken place. At the beginning, number of rows in the dataset was 1032 including complete rows and rows with missing data. 32 rows were removed because they have missing values in one or more of their features. The number of complete rows after the cleaning process was 1000. After the removal of rows with missing data, the data in all complete rows was coded. The questionnaire has a collection of questions and answers options such as numbers, text, and bins which represent some ranges such as: Age is between 31 and 40. All bins were coded to numbers in order to simplify and fast the process of feature selection. For each feature, values of codes were varied between 1 and the number of possible answers. In the second phase, rough set feature selection technique has been applied to the dataset in order to remove the redundant and irrelevant features by calculating the coefficient of each feature in the data set.

The data resulted from the preprocessing step was stored in an excel sheet and it was called a customer data set. The customer data set consists of 34 factors (questions) and 1000 complete rows (after deleting rows with noise). Factor number 34 was considered as a decision factor which reflects the customer's choice of online shopping or not. Other factors (1-33) were considered as conditional factors. The reduct of the customer data set was (Q2, Q4, Q5, Q18, Q22, Q26, Q28, and Q31) plus the decision feature (Q34). Those questions are the most important ones that affect the decision of the investors.

4. RESULTS

The first important result was shown by the generated reduct. Instead of using thirty four different features to classify a new object, only eight of them were chosen as the most significant ones. Rough set theory is pioneer in minimizing the dimensionality of datasets.

The data mining classification model for online shopping in Kuwait was built from the reduct. It is able to classify objects as online shopping preference or not. Decomposition tree, LTF-C, and Decision rules are some of the well-known classification techniques and they were used in this work. The accuracy and coverage of each classifier were calculated. High priority was given to the accuracy. The model of higher

accuracy would be used in the future to predict any unknown case. If such case does not like online shopping then companies could take steps to involve such customer in the online shopping business.

Decomposition tree builds a tree of nodes, branches, and leaves. Nodes represent conditions that check the value of the feature. Branches represent the highways which stop you at the classes. Leaves represent the classes themselves. It considered as a well-known classifier. Local Transfer Function Classifier (LTF-C) is a classification-oriented artificial neural network model. It is one of the suitable choices to classify objects. Decision rules classification is based on the relationship between conditional values and some prediction. It is good choice to classify objects.

Comparisons between the three mentioned classifiers were made using the two criterions: accuracy and coverage. The accuracy is the percentage of the instances which is classified correctly by the classifier whereas the coverage is the ratio of recognized or classified objects from the class to the number of all objects in that class. If the accuracy is the same for two classifiers then the coverage is considered as the judge. The model with higher coverage is chosen.

The accuracy and the coverage of each classifier are shown in Table 1. Decomposition tree and Decision rules are the best with full accuracy (100%) whereas LTF-C is of lower accuracy of 74.6%. The coverage of LTF-C and the Decision rules are the same and both are of full coverage (100%) whereas the coverage of the Decomposition tree is 97.2%. It is clear from the analysis results that feature selection has an impact on the classification accuracy. Based on the previous condition mentioned in the previous paragraph, the best choice will be for Decision rules, then Decomposition tree. LTF-C has no place for this dataset because its accuracy is very low comparing with other models. The model is able to predict the behavior of the new customers correctly and then the company can cluster them based on their preference as online shopping, walk to store, or both. Then the company could take a series of steps to encourage those customers who prefer online shopping by, for example, including them in their mailing list. Attract other clusters that are not preferred the online shopping by some kind of advertisements, offers, gifts, and other ways.

Table 1. The accuracy and the coverage of each model

The Model	Accuracy (%)	Coverage (%)
Decomposition tree	100	97.2
LTF-C	74.6	100
Decision rules	100	100

Figure 1 illustrates the comparison between the tested models and shows the high impact of decision rules. The figure states that decision rules and decomposition tree algorithms are able to construct classifiers with higher predictive rate than LTF-C algorithm for the current classification problem. The different in accuracy is 25.4% which shows superiority of decision rules and decomposition tree algorithms and this indeed goes along with previous findings in other business domains. Figure 1 also shows that the coverage of decision rules is higher than the coverage of decomposition tree. The difference in coverage is 2.8% for the advantage of decision rules. This yield that the decision rules is the perfect algorithm for the dataset used in this research.

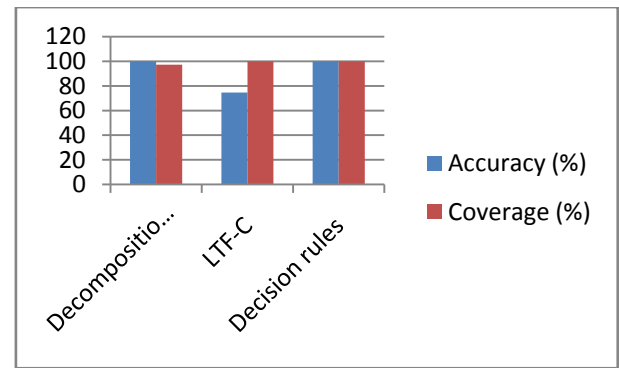


Fig 1: The comparison between the three models

Analysis of each feature in the reduct was conducted. Table 2 shows these features (questions). 746 customers out of 1000 (74.6%) preferred online shopping over walking to store where as 254 of the sample set (25.4%) preferred to walk to store and buy their items. The most important for any online shopping company is to know the kind of customers who prefer to shop online and others who prefer to visit the store. In the first case, the company works on how to encourage them to continue shop online while for the second category, the target is how to convince them to go for online shopping.

Table 2. The significant questions that represent the reduct

Question #	The question
Q2	What is your age?
Q4	What is your occupation?
Q5	Have you ever used online shopping?
Q18	How comfortable are you buying products online?
Q22	Have you had dissatisfying experience with a website which causes you to find another vendor's website? What is the reason?
Q26	What is the most important information for you when you want to buy a product?
Q28	In how many minutes do you give up searching the product (in minutes)?
Q31	How confidence do you feel when you send your credit card number over the web?

Each conditional attribute in the reduct has a strong coefficient with the decision attribute (Q34). Tables 3-10 show the relationship between each question in the reduct and the decision question. Deep navigation inside each table was taken place in order to highlight the kind of customers who preferred online shopping and the others who did not.

The category with less than 100 votes (number of customers) will not be consider as an important one since its percentage is less than 10%. In this research, threshold value is 100 (number of votes). So, in all the tables 3-10, the focus will only be on the numbers of votes that are more than or equal to the threshold value and they will be highlighted. Those votes are covering the most important categories and are of interest to the organization managements. This focus will help the management to improve and attract such categories in order to increase their profit. Other votes (less than 100) are still of interests but they are not consider in this analysis because they are of less value.

Table 3. Question 2 results

Q2	Options	Prefer to buy online	Prefer not to buy online
	11-20	58	7
	21-30	31	49
	31-40	216	109
	41-50	106	54
	>50	45	35

In Table 3, it is noticed that the customers who preferred online shopping are mostly belonging to the following criteria ($21 \leq \text{Age} \leq 50$). Votes in this criterion are 643 representing 64.3% of the whole sample. They were distributed between the following categories: 321 for ($21 \leq \text{Age} \leq 30$), 216 for ($31 \leq \text{Age} \leq 40$), and 106 for ($41 \leq \text{Age} \leq 50$). 109 members (10.9%) of the sample who did not prefer online shopping are of age between 31 and 40 ($31 \leq \text{Age} \leq 40$). If it is assumed that age between 21 and 50 represents young and mid-aged then it is concluded that this category is the most important for the success of the web-based selling. Companies should focus on the needs of this category. On the other hand, companies must not ignore the other category ($31 \leq \text{Age} \leq 40$) which represents 10.9% of the sample and they must find ways to attract this category and involve it in the web-based selling business. The comparison of question 2 results is shown in Figure 2.

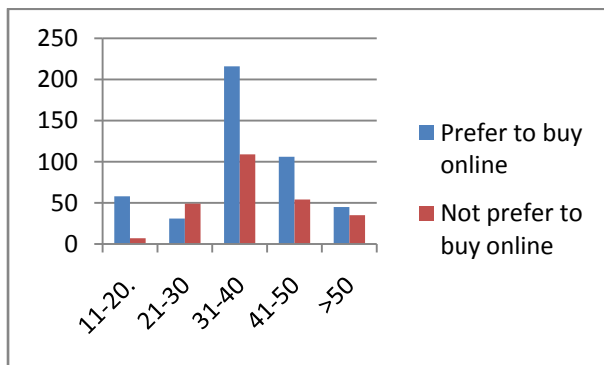


Fig 2: The comparison of question 2 results

Table 4. Question 4 results

Q4	Options	Prefer to buy online	Prefer not to buy online
	Self-employees	60	15
	Employee	315	49
	Professional	184	63
	Student	117	67
	Retired	50	23
	Others	20	23

In Table 4, it is noticed that the customers who preferred online shopping are mostly belonging to the following categories: employee, professional employee, and student. Votes for each category are 315, 184, and 117 respectively and they are all equal to 616 which represents (61.6%) of the whole sample. None of the sample members who did not

prefer online shopping is deliberated since the votes in each category is less than the threshold value. Companies should focus on the needs of those mentioned categories. On the other hand, companies must find some ways to attract other non-interested categories. The comparison of question 4 results is shown in Figure 3.

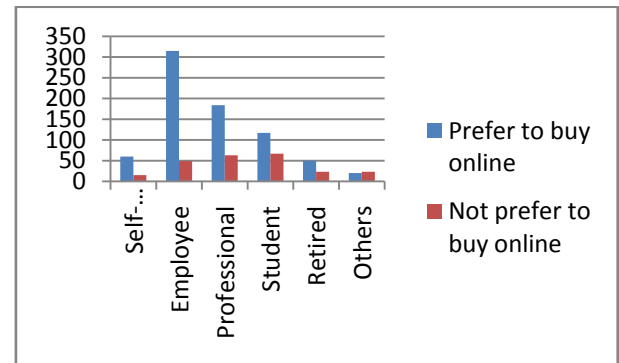


Fig 3: The comparison of question 4 results

Table 5. Question 5 results

Q5	Options	Prefer to buy online	Prefer not to buy online
	Yes	535	124
	No	211	130

In Table 5, it is noticed that the two categories shared interested and non-interested customers. 535 members of the whole sample who preferred to buy online are already using the online shopping, whereas 211 are not. On the other hand, 124 members of the whole sample who did not prefer to buy online are using the online shopping whereas 130 are not. It is concluded that 65.9% of the sample used the online shopping. 53.5% are happy of using the online shopping and they adopt it as their main shopping way. 12.4% used the online shopping but they are not satisfied of it since they do not prefer this way of shopping. 21.1% did not use the online shopping but they are interested to start with this way. Most probably they don't have computer, internet, credit card, and/or other possibilities. 13% are not using the online shopping and not prefer to buy online as well. Companies may focus more on improving their online shopping website (the interface, the flexibility, security, offers, and so on) in order to encourage the 12.4% to contribute in the online shopping systems. Also, companies should make deep investigation on how to involve other 21.1% and 13% categories in the online shopping business. The comparison of question 5 results is shown in Figure 4.

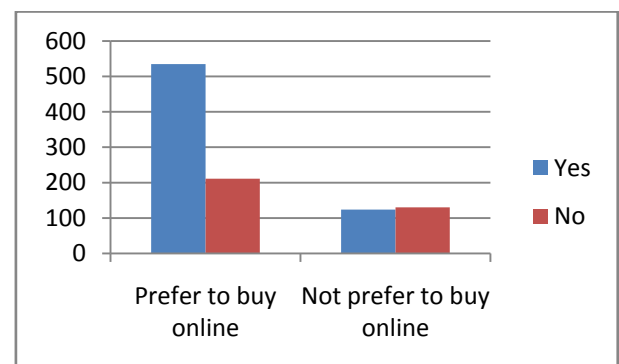


Fig 4: The comparison of question 5 results

Table 6. Question 18 results

Q18		
Options	Prefer to buy online	Prefer not to buy online
Extremely	79	36
Quite	184	39
Moderately	289	81
Slightly	182	68
Not Comfortable	48	30

In Table 6, it is noticed that the customers who preferred online shopping are mostly belonging to the following categories: quite, moderately, and slightly. Votes for each category are 184 (represents 18.4%), 289 (represents 28.9%), and 182 (represents 18.2%) respectively and they are all equal to 655 votes which represents (65.5%) of the whole sample. None of the sample members who did not prefer online shopping is deliberated since the votes in each category is less than the threshold value. Companies should focus on the needs of those mentioned categories. The website of the online shopping should be improved in order to convince the mentioned categories to use the online shopping website comfortably so we will not lose them in the future. On the other hand, companies must find some ways to attract other non-interested categories. The comparison of question 18 results is shown in Figure 5.

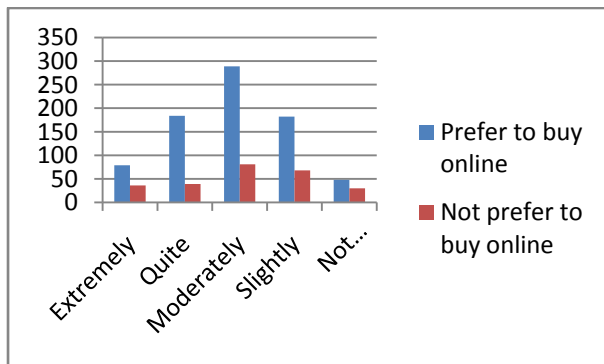


Fig 5: The comparison of question 18 results

Table 7. Question 22 results

Q22		
Options	Prefer to buy online	Prefer not to buy online
No	105	27
Yes, confusing	189	57
Yes, poor pages	269	81
Yes, annoying	147	60
Yes, but slow downloading	36	29

In Table 7, it is noticed that the number of the customers above or equal the threshold value (100) is 710. They preferred online shopping and they represent 71%. 10.5% of them satisfy with the website whereas 60.5% have some problems with the website such as: confusing with the website (18.9%), the poor design of the website (26.9%), and the

annoying website (14.7%). They still prefer online shopping even they have such bad experience with the website as mentioned above. There are no interesting figures about those who did not prefer to buy online. The comparison of question 22 results is shown in Figure 6.

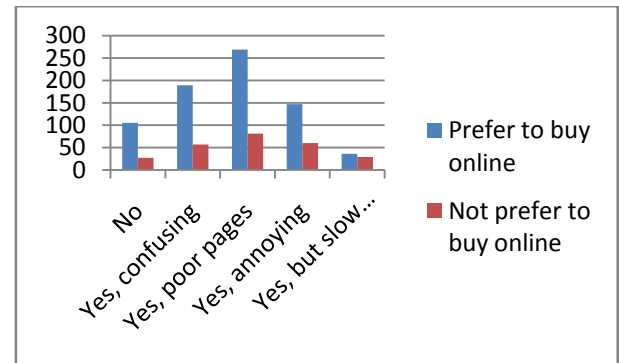


Fig 6: The comparison of question 22 results

Table 8. Question 26 results

Q26		
Options	Prefer to buy online	Prefer not to buy online
Known store	283	79
Product info	65	17
Price	155	50
Availability	181	73
None	62	35

In Table 8, 61.9% of the whole sample over or equal to the threshold value preferred to buy online. Well-known store, price of the item, and availability of the item are the most interest things for the customer who wants to buy online. There are no interesting figures shown about customers who preferred not to buy online. This is a big challenge and the companies must study how to attract and involve them in the online shopping family. The comparison of question 26 results is shown in Figure 7.

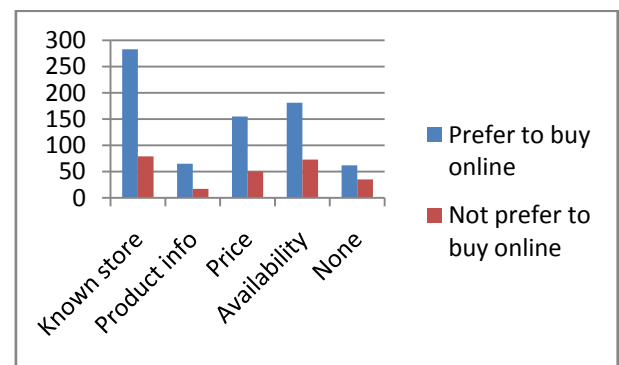


Fig 7: The comparison of question 26 results

Table 9. Question 28 results

Q28		
Options	Prefer to buy online	Prefer not to buy online
<5	81	15
5-10	108	22
11-20	183	47
21-30	176	56
31-60	96	52
>60	68	38
Never noticed	34	24

In Table 9, 467 member of the sample (46.7%) who preferred to buy online are giving up searching the item in 5-30 minutes. This is an attractive result which encourages the management of the online shopping website to improve it and make the searching process faster. Also they may look carefully at their server specification and the structure of the website design and programming in order to develop its slow processing. There are no interesting figures about those who did not prefer to buy online. The comparison of question 28 results is shown in Figure 8.

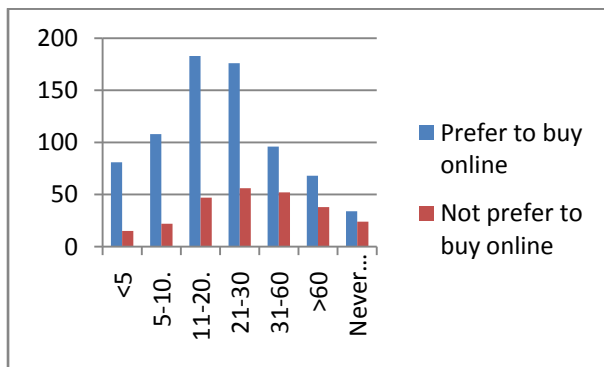


Fig 8: The comparison of question 28 results

Table 10. Question 31 results

Q31		
Options	Prefer to buy online	Prefer not to buy online
Close to 100%	65	14
Close to 75%	165	56
Close to 50%	294	81
Close to 25%	170	66
Close to 0%	52	37

Table 10 illustrates those customers (over or equal to the threshold value) who preferred to buy online and they accept to send their credit cards numbers on the web with confidence between 25%-75%. Their percentage is 62.9%. 16.5% send their credit card number over the web with confidence 75%. 29.4% send their credit card number over the web with confidence 50%. 17% send their credit card number over the web with confidence 25%. The companies should have ambition to reach with them to 100% confidence. To do so,

they must guarantee the security of their websites and avoid any fake. There are no interesting figures about those who did not prefer to buy online. The comparison of question 31 results is shown in Figure 9.

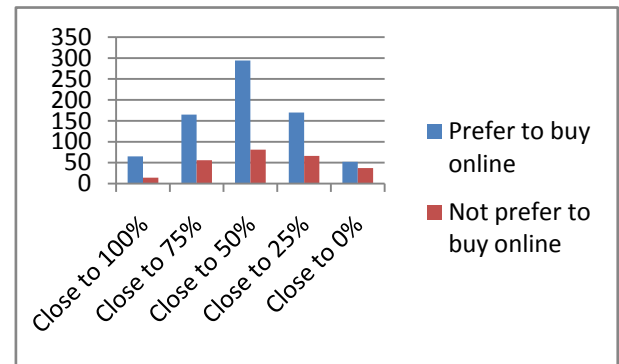


Fig 9: The comparison of question 31 results

Figures 1, 3, 4, 5, 6, 7, and 8 show that customers who preferred to buy online are always more than those who did not except for one non-significant case which is the category "others" that belongs to Q4. The difference is only 3 votes.

5. CONCLUSION

Following conclusion may be drawn from the data mining classification model obtained using rough set approach. It has been found that there are three major obstacles in the growth of online shopping market in Kuwait, which are as follows:

1. Customers are not comfortable using online shopping (A culture of fear of what's new).
2. Kuwait markets lack supporting online shopping. This is clearly shown by poor website design, confusing website, and annoying website, relatively long response time.
3. Customers fear of using their credit cards on the web. They do not trust the online market.

From the data mining model and the features reduction analysis results, it is clear that feature selection has an impact on the classification accuracy. Eight features out of the thirty three predicting features are able to predict new example correctly by decision rules and decomposition tree algorithms with an advantage for decision rules since it has slightly better coverage percentage. This proves that the selected features (reduct) are the most important predicting features for the success online shopping system. The features were shown by Q2, Q4, Q5, Q18, Q22, Q26, Q28, and Q31.

This article indicates that the development of online shopping plays effect to online shopping growth in Kuwait. Online shopping companies have to improve their online shopping websites, work hard to attract customers to use the online shopping, increasing the effort to encourage using the online shopping, and improve the self-confidence of customers to use their online shopping websites.

Increasing the investment in online shopping business will increase the competition between similar companies. This will then play a role to improve the infrastructures and resolve the drawbacks of current online shopping environment.

In near future, I will explore the importance of other feature selection methods and study their effect on improving the online shopping system in Kuwait. Also, I will study the

effect of the categories that have less than 100 votes (number of customers) on the profit's growth of MLShCs in Kuwait.

6. ACKNOWLEDGMENTS

This research was funded by the Arab Open University/Kuwait Branch. The author likes to thank many anonymous people for their efforts in improving the readability of this paper including the patience of my wife and kids.

7. REFERENCES

- [1] WTO. 1998. Study from WTO Secretariat highlights potential trade gains from electronic commerce. <http://www.wto.org/english/newse/pres98e/pr96e.htm>
- [2] ECLAC. 2002. Electronic Commerce, International Trade and Employment: Review of the Issues. UN, Economic commission for Latin America and the Caribbean ECLAS, Washington Office, April, pp: 1-30.
- [3] Yu, Z., Jing, B., and Weixiang, Z. 2013. Trust fraud: A crucial challenge for China's e-commerce market. *Electronic Commerce Research and Applications*, 12: 299-308.
- [4] Ramakrishnan, R., Usha, R., and Hsieh-Ling, H. 2012. The impact of e-commerce on Taiwanese SMEs: Marketing and operations effects. *International Journal of Production Economics*, 140: 934-943.
- [5] Panagariya, A. E-Commerce, WTO, and Developing Countries. 2000. Policy issues in international trade and commodities study Series. No.2 UN, New York and Geneva, pp: 1-33.
- [6] Malkawi, B. H. 2007. E-commerce in Light of International Trade Agreements: The WTO and the United States-Jordan Free Trade Agreement. *International Journal of Law and Information Technology*, 15: 153-169.
- [7] Nuray Terzi, 2011. The impact of e-commerce on international trade and employment. *Procedia - Social and Behavioral Sciences*, 24: 745-753.
- [8] Dave C. 2011. E-Business and E-Commerce Management: Strategy, Implementation and Practice, 5th Edition, Prentice Hall, New York.
- [9] Jarvenpaa, S. L., Tractinsky, N., and Vitale, M. 2000. Consumer Trust in an Internet Store. *Information Technology and Management*, 1: 45-71.
- [10] Kim, J. and Moon, J. Y. 1998. Designing Emotional Usability in Customer Interfaces – Trustworthiness of Cyber-banking System Interfaces. *Interacting with Computers*, 10: 1-29.
- [11] Egger, F. N. 2000. Trust Me, I'm an Online Vendor: Towards a Model of Trust for E-Commerce System Design. in: G. Szwillus & T. Turner (Eds.): CHI2000 Extended Abstracts: Conference on Human Factors in Computing Systems, The Hague (NL), ACM Press, April 1-6, pp: 101-102.
- [12] Inc. Magazine - December 2010/January 2011 English.
- [13] Parasuraman, A., Zeithaml, V. A. and Berry, L. L. 1988. SERVQUAL: a multiple item scale for measuring customer perceptions of service quality. *Journal of Retailing*, 64: 12-40.
- [14] Kuo, Y. F. 2003. A study on service quality of virtual community web sites. *Total Quality Management*, 14: 461-73.
- [15] Negash, S., Ryan, T. and Igbaria, M. 2003. Quality and effectiveness in web-based customer support systems. *Information and Management*, 40: 757-68.
- [16] Al-Hudhaif, S. and Alkubeyyer, A. 2011. E-Commerce Adoption Factors in Saudi Arabia. *International Journal of Business and Management*, 6: 122-133
- [17] Oreku, G. S., Mtenzi, F. J. and Ali, A. D. 2011. The Prospects and Barriers of E-Commerce Implementation in Tanzania. *Conference Proceedings, ICIT 5th International Conference on Information Technology*, Amman, Jordan, pp: 11-13.
- [18] Radovilsky, Z. and Hegde, V. G. 2004. Factors influencing e-commerce implementation: Analysis of survey results. *Journal of Academy of Business and Economics*, 4: 29-37
- [19] Zhu, K. 2004. E-Commerce Capability: A Resource-Based Assessment of Their Business Value. *Journal of Management Information Systems*, 21: 167-202.
- [20] Al-Fadhli, S. 2011. Critical Success Factors influencing e-commerce in Kuwait. *Journal of Internet Banking and Commerce*, 16.
- [21] Pawlak Z. and Skowron A. 2007. Rudiments of rough sets. *Information Sciences*, 177: 3-27.
- [22] Al-Shalabi L. 2009. Improving Accuracy and Coverage of Data Mining Systems that are built from Noisy Datasets: A new Approach. *Journal of Computer Science*, 5: 131-135.
- [23] Al-Shalabi L. 2016. Data Mining Application: Predicting Students' performance of ITC program in the Arab Open University in Kuwait – The blended Learning. *International Journal of Computer Science and Information Security (IJCSIS)*, ISSN 1947-5500, Pittsburgh, PA, USA, 14: 827-833.
- [24] Al-Shalabi L. 2017. Perceptions of crime behavior and Relationships: Rough Set Based Approach. *International Journal of Computer Science and Information Security (IJCSIS)*, ISSN 1947-5500, Pittsburgh, PA, USA, 15: 413-420.
- [25] Pyle, D. 1999. *Data Preparation for Data Mining*. Morgan Kaufmann Publishers, LosAltos, California.
- [26] Al-Shalabi, L., Shaaban, Z., and Kasasbeh, B., 2006. Data mining: A Preprocessing Engine. *Journal of Computer Science*, 2: 735-739.
- [27] Quinlan J. R. 1989. Unknown Attribute Values in Induction. In *Proceedings of the Sixth International Workshop on Machine Learning*, pp: 164-168.
- [28] Dempster, A. P., Larid, N. M., and Rubin, D. B. 1977. Maximum likelihood from incomplete data via the Em Algorithm (with discussion). *Journal of Royal Statistical Society. B39: 1-38*.