Explaining Individuals' Usage of Social Commerce: A Data Mining Approach

Yazn Alshamaila[1], Ibrahim Aljarah[1] & Ala' M. Al-Zoubi[1]

[1] King Abdullah II School for Information Technology, The University of Jordan, Amman, Jordan

Correspondence: Yazan Alshamaileh, The University of Jordan, Business Information Technology, Amman, Jordan. E-mail: Y.Shamaileh@ju.edu.jo

Abstract

With the use of Web 2.0 technology, e-commerce is undergoing a radical change that enriches consumer involvement and enables a better understanding of economic value. This emerging phenomenon is known as social commerce. Social commerce (s-commerce) presents a new alternative for consumers to search for and find information about products they are seeking to buy. In spite of its universality, the adoption of this burgeoning technology is affected by several factors. This research project is an initial attempt to explore individuals' intention of s-commerce usage through the data mining approach. The data was collected via a web-based questionnaire survey of 360 social network site (SNS) users in Jordan. Data mining techniques were then used to analyze the collected data in order to figure out what group of features is best for predicting s-commerce adoption among SNS users. The results showed that data characteristics related to gender, monthly income, civil status, number of connections, and prior online shopping experience are key factors in the classification process. The findings may assist researchers in investigating social commerce issues and aid practitioners in developing new s-commerce strategies.

Keywords: social commerce, social network sites, data mining

1. Introduction

In today's rapidly evolving online world, the growth of social network sites is continually bringing new concepts to light. One of the main concepts that has emerged due to the existence of both Web 2.0 technologies and social media is social commerce (s-commerce). According to Zeng, Huang, and Dou (2009), online transactions (e-commerce) have developed into s-commerce due to the interactions and connections between people, especially on social networking sites (SNSs). S-commerce differs from e-commerce in that it is established on several kinds of social media platforms, including Facebook, Instagram, and Twitter. It uses social media to support commercial activities (Chen & Shen, 2015). S-commerce is often represented as a subarea derived from e-commerce (Hajli, 2013; Sturiale & Scuderi, 2013); Wu, Shen, and Chang (2015) defined it as "word-of-mouth applied to e-commerce."

Although social commerce is a new trend, this phenomenon has developed quickly and remarkably (Barnes, 2014). Social platforms such as social networking sites have their own role in the s-commerce advancement. A report conducted by Barclays (2012) shows that half of the U.K. consumer population will be part of the s-commerce phenomenon by 2021. A new report conducted by We Are Social reveals that 2.80 billion users are active on various social network sites, amounting to 37% of the total global population in 2018 (Kemp, 2018). HnyB Insights (2014) published a report stating that the revenue of global s-commerce is growing year by year, predicting the market to reach $80 billion U.S. dollars by 2020. Popular social networking sites such as Facebook, Instagram, and Twitter have their own impact on s-commerce growth. Facebook is considered to have a noticeable share in s-commerce activities. In 2015, the number of purchases increased to 52%, in comparison with the previous year (McCarthy, 2015). According to Chaykowski (2015), 50 million small companies have Facebook pages to communicate with their customers, 4 million companies pay for advertisements on Facebook, thus demonstrating the platform's influence on electronic commerce in general and s-commerce in particular. Facebook's Instagram subsidiary also plays an important role in s-commerce; on that platform, 4.21% of users interact with popular commercial brands, a rate 58 times higher than brand engagement on Facebook (Mathison, 2018).

A growing body of literature recognizes the importance of s-commerce and SNS usage among people of different ages and social statuses. In this context, the greatest significance of s-commerce is its capacity to become the most universally adopted platform for electronic commerce worldwide (Zhang & Benyoucef, 2016). To date, however, there are few studies published in this field. Nevertheless, authors of studies such as the one conducted by Chen and Shen (2015) assert that research on s-commerce is increasing. To take advantage of s-commerce, Hennig-Thurau, Hofacker, and Bloching (2013) and Liébana-Cabanillas and Alonso-Dos-Santos (2017) believe that s-commerce consumer behavior should be studied. In the Hashemite Kingdom of Jordan, 67% of adults have internet access, whereas 90% of these users are active on SNSs; in comparison, rates in other economically advanced countries such as the United States and the United Kingdom are lower: 71% and 66%, respectively (Poushter, 2016). Thus, this issue is considered vital in this setting and area, and it requires an extensive push for further study. In an attempt to achieve this goal, the key objective of this paper is to examine patterns in the existing data to predict the factors that may affect s-commerce adoption depending on personal and SNS usage-related features.

The paper is divided into five sections. In the introduction, we present the logic and main purpose of the study; in the second section, we review previous studies in this field. The third section consists of the research methodology, followed by a comparative analysis and its results in the fourth section. The paper will end with a summary of the study's findings and a discussion on possible future research.

2. Background

S-commerce, or social commerce, is defined as commerce activities mediated through social network sites. Jascanu, Jascanu, and Nicolau (2007) stated that through different SNS avenues such as reviews, ratings, chat rooms, locator applications (geo-tagging), ranking, and recommendations, customers have been enabled to share their information, experiences, and opinions related to different products and services. Yahoo! was the first to introduce the term "social commerce" in 2005 (Jascanu et al., 2007), and since that time, it has been described as a new and powerful phenomenon. This novel advancement in e-commerce makes consumers active participants in the online business process (Hajli, 2015). Statistics show that s-commerce has developed notably since its inception, allowing the creation of new business models based on online communities (Hajli, 2015). Social commerce has gained wide popularity since 2005. For example, a survey with 2 000 respondents conducted by Immediate Future (2010) indicated that 53% of U.K. consumers review products and services online. This survey also revealed that those reviews are 157% more influential than traditional advertising methods. A report conducted by McKinsey revealed that Chinese people spend 78 minutes per day in s-commerce activities (as cited in Chiu, Ip, & Silv, 2012), and around 50% percent of Chinese people buy products depending on the online recommendations of relatives and friends (Liu, Chu, Huang, & Chen, 2016).

S-commerce has benefits for both customers and firms because s-commerce platforms simplify both consumer-to-consumer (C2C) and business-to-consumer (B2C) connections. According to Xiang, Zheng, Lee, and Zhao (2016), consumers have the ability to interact with other consumers who can help them when choosing products and services. Moreover, they can take advantage of ratings and reviews of products and sellers left by other online actors (Zheng, Zhu, & Lin, 2013). Based on these reviews and ratings, customers can evaluate a product's quality and decide whether they desire it. Additionally, they may know more about the seller's credibility and the online shopping experience. Jascanu et al. (2007) state that social commerce gives consumers an opportunity to share information and express their opinions about purchasing products. Kietzmann, Hermkens, McCarthy, and Silvestre (2011) argue that social media influences a firm's reputation and impacts its transactions. As a result, firms should pay attention to social media and the emergence of new concepts such as s-commerce.

Several previous studies indicate that s-commerce has a positive influence on firms. Kim and Park (2013) state that 300 Korean s-commerce firms have gained approximately $300–$500 million in sales. Thus, firms can increase their profits by attracting buyers via positive recommendations (Curty & Zhang, 2011). To take advantage of this form of commerce, firms may use SNSs in an attempt to sell products, in addition to asking their consumers to comment and write about their experiences on social platforms in order to introduce their products and services. Via strategic use of s-commerce, companies can enhance their relationships with consumers and therefore increase both sales and brand loyalty (Hajli, 2014). Moreover, s-commerce provides businesses with a variety of ways to build customer relationships, reduce marketing expenses, and increase sales (Chen, Lu, & Wang, 2017). Dell and H&M are examples of major companies that use s-commerce to present their products (Hajli, 2015).

However, several factors influence the presence of customers in s-commerce. Wigand, Benjamin, and Birkland

(2008) state that customers' need to be independent, successful, and connected to others are key factors that motivate them to participate in s-commerce. Sharma and Crossler (2014) use both the trust theory and uses and gratifications theory to clarify consumers' intention to adopt s-commerce. They found that trust factors such as security, privacy, and information quality influence the intention to be engaged in s-commerce. Zhang, Gupta, and Zhao (2014) designed a model based on a stimulus–organism–response model to investigate the factors affecting participation in s-commerce. Technological environmental features such as interactivity, personalization, and sociability have an impact on the customer's experience regarding social support, social presence, and flow, which in turn influence the intention of s-commerce adoption. Thus, interactions between customers (i.e., expressing their opinions and sharing information) will increase the trust and therefore increase the social support.

Previous studies have established that gender, age, prior online shopping experiences, and income affect consumers' adoption of e-commerce and s-commerce. Rodgers and Harris (2003) studied the role of gender in e-commerce adoption, which was analyzed together with the role of prior experience of online shopping. Other studies such as those by Zhang, Benyoucef, and Zhao (2015); Rapp, Beitelspacher, Grewal, and Hughes (2013); Z. Li and C. Li (2014); and Chang, Yu, and Lu (2015) hold that consumer demographics such as age and income play a role in s-commerce adoption in addition to being factors that may help companies and industries leverage their marketing potential.

3. Research Methodology

This research applies data mining techniques and methods to generate new knowledge using data collected from a launched survey about social commerce. One of the aims of this research is to examine the capacity of data mining techniques to predict s-commerce usage among SNS users. The study uses a mixed-methods approach based on three steps. Namely, examine datasets against main classifiers; feature selection and weight the features using decision tree. The implementation of this study depends on WEKA software, which offers several classification methods (Witten, Frank, Hall, & Pal, 2016).

The study survey contained general socio-demographic questions such as age, gender, income, and study and work field. The remaining sections were related to technology use such as type of smartphone used, internet access, experience in the internet field, and experience with s-commerce and e-commerce. In total, 49 features were extracted. The collected data were analyzed in order to discover any common features between heavy and non-heavy s-commerce users.

Data were classified depending on how many times the respondent logs on to social networks to make purchasing decisions about a new product or service. The classes are split into six categories: None, 0–2, 3–5, 6–8, and 9–11/times per week, and more than 12 times per week. In this project, two data sets were generated from the same collected data.

Table 1. Dataset2 – Key

Features	Key	Count	percentage
F1 Gender			
Female	1	218	61%
Male	2	142	39%
		360	100%
F3 Civil status			
Single	1	205	57 %
Married	2	155	43%
		360	100%
F6 Monthly income			
Less than 150 JD	1	78	21%
150 -300 JD	2	53	15%
300-500 JD	3	65	18%
500-800 JD	4	46	13%
800-1000JD	5	42	12%
1000-2000 JD	6	22	6%
2000-3000 JD	7	40	11%
More than 3000 JD	8	14	4%

			360	100%
F37 Number of connections on SNS				
Less than 50	1		37	10%
100-50	2		46	13%
300-100	3		102	28%
500-300	4		60	17%
1000-500	5		60	17%
2000-1000	6		22	6%
3000-2000	7		4	1%
More than 3000	8		29	8%
			360	100%
F49 Prior experience – online shopping				
Yes	1		180	50%
No	0		180	50%
			360	100%
Class -- how many times using social commerce per week				
(0-2)	1		190	53%
(3-5)	2		0	0%
(6-8)	3		43	12%
(9-11)	4		14	4%
More than 12 times	5		54	15%
I don't use online shopping	6		59	16%
		(Low: 1+6) (High:3+4+5)	360	100%

In the first data set (Dataset1), we gathered the 0–2 and None categories, classifying them as the Low class, whereas the rest of the categories (3–5, 6–8, 9–11, more than 12) were classified as High. However, after running an analysis, the accuracy results obtained from Dataset1 were 0.68. In an attempt to make clear separation between the two classes (high and low class), and to get better accuracy results, we modified our classification from Dataset1 to generate Dataset2 as follows: we deleted the 3–5 category in order to clearly distinguish between the heavy s-commerce users' group and the low s-commerce users' group. Then we gathered the 0–2 and None categories and classified them as a Low class, whereas the 6–8, 9–11, and more than 12 categories were gathered under the High class (Table 1). The sample evaluated in Dataset1 contained 452 full responses. After modification (i.e., deleting the 3–5 category), in total, 360 full responses remained in Dataset2; of these, 92 cases were deleted.

Our analysis proceeded along the following steps. First, we examined the two datasets against well-known classifiers, which we will explore in the next section. In the second step, the best subset of features was selected by applying an auto feature selection process. Finally, in the third step, we studied the weight of each feature by using the decision tree algorithm. We illustrate more details about the analysis process in the following section.

4. Experiments and Results

4.1 Data Mining Methods

Data mining uncovers new significant connections, patterns, and orientations by analyzing the data through statistics, machine learning, artificial intelligence (AI), and data visualization techniques. This process efficiently points out implicit, previously obscure, and potentially useful information found in the collected data, thus enabling the discovery of predictive patterns, the creation and testing of hypotheses, and the production of insight-provoking visualizations (Han, Pei, & Kamber, 2011).

4.1.1 Classification Algorithms

In the first step of data analysis, we selected several classification algorithms that had the potential to yield good results. Some well-known WEKA classifiers were used in addition to two decision tree algorithms (J48 and REPTree), the Naive Bayes (NB) Bayesian classifier, Multilayer Perceptron (MLP), and a Nearest Neighbor algorithm (k-NN).

Reduced Error Pruning Tree (REPTree)

A classifier used in dealing with noise in decision tree learning, the REPTree was introduced in the post-pruning process based on the ideas of Quinlan (1987) and Pagallo and Haussler (1990; Brunk & Pazzani, 1991). Fürnkranz

and Widmer (1994) described the algorithmic process as follows: in the first step, the data are divided into two sets. The first is the growing set, generated through a realization of learning algorithms, whereas the second is the pruning set, which is produced through deleting literals and clauses from the theory until it results in a predictive accuracy decrease measured through the pruning set.

J48 Decision Tree

J48 is a decision tree stemming from the C4.5 decision tree algorithm (Kabakchieva, 2013). The C4.5 decision tree algorithm points at data through information entropy (Quinlan, 1993). Al-Zoubi, Alqatawna, and Faris (2017) described the repeated process applied on smaller subsets as follows: an attribute is selected from the collected data at each node by splitting the representative into subsets through the information gained criteria, resulting in a decision based on the highest attribute value.

Naive Bayes (NB)

The Naive Bayes (NB) is a probability theory-based algorithm that receives an independent contribution from each feature and exports it to the output class (al-Zoubi et al., 2017). In this process, no explicit classifiers are used. Despite its simplicity, Han et al. (2011) defended the algorithm's equality of accuracy in pragmatic applications.

Multilayer Perceptron (MLP)

Multilayer Perceptron (MLP) is one of the more commonly approached artificial neural networks; it uses one or multiple hidden layers embedded as an aid to help expand the network's ability to model complex functions (Paola & Schowengerdt, 1995). Al-Zoubi et al. (2017) described its function as an information processing system consisting of several layers used to map the input data into fitting sets of outputs. In the MLP, the input layer is passive, whereas both the hidden and output layers process the data actively (Zare, Pourghasemi, Vafakhah, & Pradhan, 2013).

k-Nearest Neighbor (k-NN)

The k-Nearest Neighbor algorithm (k-NN) is a closest training example in feature space classifier that is based on instance learning or lazy learning (Aha, Kibler, & Albert, 2013). The close instance is determined by using distance measurements and functions such as Euclidean, Minkowski, and Mini- max (Al-Zoubi et al., 2017). Kabakchieva (2013) described it as one of the most average machine learning algorithms because its classification process is established on a majority vote from the object's neighbors, thus assigning it to the most common class amid the k-nearest neighbors.

4.2 Evaluation Measurements

Figure 1, and table 2 shows the evaluation criteria used in this work, which is known as a confusion matrix.

		Predicted class	
		Positive	Negative
Actual class	Positive	True Positive (TP)	False Negative (FN)
	Negative	False Positive (FP)	True Negative (TN)

Figure 1. Confusion Matrix

Table 2. Evaluation Criteria Description

Evaluation criteria	Equation	Description
Accuracy	$Accuracy = \dfrac{TP + TN}{TP + FN + FP + TN}$	Estimated by dividing the numbers of all correct instances by that of all instances in the data.
Precision	$Precision = \dfrac{TP}{TP + FP}$	Estimated by the division of the percentage of correctly predicted instances by the total number of

		predicted instances.
Recall	$$Recall = \frac{TP}{TP + FN}$$	Calculated by dividing the ratio of correctly classified instances by the number of actually classified instances.
F-measure	$$F-measure = \frac{2 \times Precision \times Recall}{Precision + Recall}$$	Calculated by the weighted average of the Recall and Precision.

We examined Dataset1 and Dataset2 against the classifiers: J48, REPTree, NB, MLP, and k-NN. As explained previously, Dataset1 contains all the categories, with (0–2 and None) classified as Low and (3–5, 6–8, 9–11, and more than 12) classified as High. Table 3 shows the best accuracy rate for Dataset1 as 60.84%, achieved by the REPTree classifier, followed by the NB and J48 with 59.96% and 57.96%, respectively. MLP and k-NN demonstrated the worst accuracy rates, with 54.87% for each. The best recall and F-Measure were also obtained by REPTree, whereas the second-best recall and F-Measure were achieved by NB, with 62.2% and 63.1% respectively. Moreover, the highest classifier for the precision was the NB, with a 64% rate, and the second highest results were achieved by REPTree.

Table 3. Dataset1 Results

Data1	All Features			
Measure	**Accuracy**	**Recall**	**Precision**	**F-Measure**
J48	57.964	0.614	0.619	0.617
RepTree	**60.840**	0.667	0.638	0.652
NB	59.955	0.622	0.640	0.631
MLP	54.867	0.582	0.592	0.587
k-NN	54.867	0.574	0.593	0.584

Unlike Dataset1, Dataset2 achieved higher results (see Table 4). By removing the 3–5 category from the main class, the results improved in all measures. The best classifier for sorting the High and Low users was REPTree, with a 66.66% accuracy rate. The second-best result was obtained using the NB classifier. However, similar to Dataset1, REPTree achieved the best results for F-Measure and recall measures, with 78.6% and 88.8%, respectively.

Table 4. Dataset2 Results

Data2	All Features			
Measure	**Accuracy**	**Recall**	**Precision**	**F-Measure**
J48	61.111	0.767	0.700	0.732
RepTree	**66.667**	0.888	0.706	0.786
NB	65.556	0.711	0.773	0.741
MLP	61.389	0.727	0.718	0.723
k-NN	61.111	0.675	0.740	0.706

As mentioned in Section 4.1, the data with the best results, Dataset2, will go through an auto feature selection process.

4.3 Feature Selection (FS)

Feature selection (FS) can be applied to increase the accuracy rate of the classification (Al-Zoubi et al., 2017). In the machine learning context, the FS method reduces the dimensionality of data without losing any information. This technique is used in the processing phase of the methodology to try to select the best subset of features and remove unneeded or irrelevant features (Faris, Ala'M, & Aljarah, 2017; Faris et al., 2018), thus improving the results.

Table 5. Dataset2 Auto Features Selection

Dataset2-best-Features	TOP-5				
Measure	Accuracy	Recall	Precision	F-Measure	Auto feature selection process-Result
J48	**69.167**	0.932	0.712	0.807	**F1:** Gender.
RepTree	68.056	0.944	0.699	0.803	**F3:** Civil states.
NB	**69.167**	0.843	0.745	0.791	**F6:** Average monthly income.
MLP	68.056	0.863	0.726	0.789	**F37:** The number of connections on the most used social network platforms.
k-NN	68.611	0.839	0.741	0.787	**F49:** Prior experience in online shopping.

Next, as a second step, we analyzed the data with the highest results using an auto feature selection method to select the best feature subset. In our case, out of 49 features collected via online survey, five features were selected using the auto feature selection technique: F1: Gender: female; F3: Civil status: single; F6: Average monthly income; F37: Number of connections on the most used social network platforms; and F49: previous experience in online shopping.

As shown in Table 5, all measures achieve a better result than the previous experiments did. The best accuracy was obtained by the J48 and NP, with a nearly 2.5% increase over all feature data in Table 3. In addition, the best recall and F-Measure also show a notable increase, with a 5.6% and 2.1% increase, respectively, which was achieved by REPTree. The classifier with the highest precision was the NB, with 74.5%.

4.4 Decision Tree

A decision tree is a tree-form graph of the possible consequences and outcomes of several choices. The tree helps to weigh these possible choices based on their probabilities, cost, and benefits. To be more specific, it is a decision support tool that can be used to map an algorithm in order to mathematically predict the most suitable choice (Safavian & Landgrebe, 1990). In our case, the next step in analysis was to run the selected features on the decision tree classifier in order to analyze and weigh the features. The outcomes are illustrated in Figure 2.
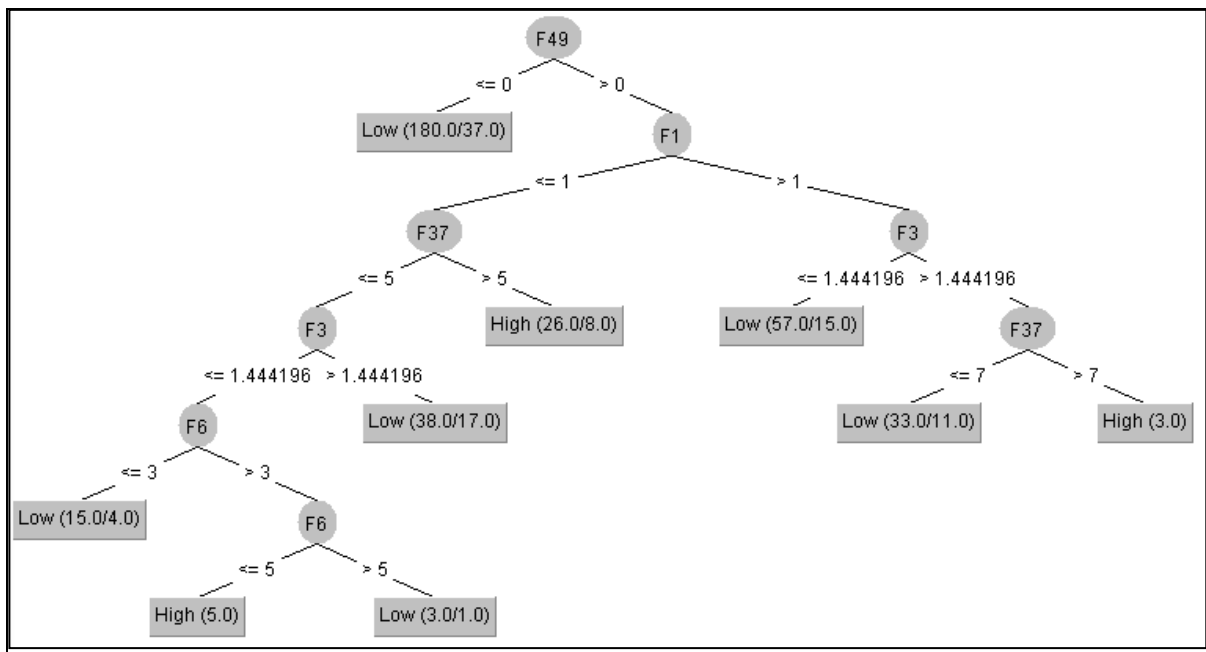

Figure 2. J48 Tree Model Dataset2

A summary of the rules that can be concluded from the J48 Tree Model Dataset2 in Figure 2 will be highlighted in the following section.

5. Discussion

As several reports have suggested, s-commerce is characterized by two facets: social sharing and social shopping (Chen & Shen, 2015). Arising from the feature selection technique, the number of users' connections on SNSs (F37) is among the best selected subset features that could influence the users' adoption of social commerce services. This finding validates the assertion that social shopping is affected by a person's relationship with the surrounding community and other consumers, especially when consumers' acquaintances take part in shopping decisions and other social practices that take place during the process (Chen & Shen, 2015). This study supports evidence from previous observations (e.g., Chen & Shen, 2015; Hajli, 2013) that demonstrated that both informational support and emotional support from a user's connections on social network can have an effect on his or her online shopping behavior.

Notably, the results confirm the relationship between gender (F1) and civil status (F3) and their association with intensive use of s-commerce. This finding supports evidence from Awad and Ragowsky's (2008) previous observations that women place greater value on online comments from other consumers, which confirms the role of gender in online shopping. Moreover, many researchers, including Campbell (2000) and Dittmar, Long, and Meek (2004), have noted that female users have a more positive attitude regarding online shopping compared to male users.

This study's findings also indicate that users with good prior experience of buying online from different websites (F49) will be strong candidates to be heavy s-commerce users. In addition, our findings support the idea that consumers who had a long and positive previous experience with shopping online using traditional websites such as Amazon and eBay will have a high capacity to use social networks to seek information for products and make an actual purchase decision. In other words, the users who have these characteristics will be strong social commerce users. Although immediate conclusions are available, these findings need further analysis

5.1 Algorithm mapping - Decision Tree Rules

Table 6 depicts a summary of the extracted rules from Figure 2. The rules encompass five features: (F1) representing gender, (F3) representing civil status, (F6) representing monthly income, (F37) representing number of connections on SNSs, and (F49) representing prior experience in online shopping.

Table 6. Rules for Dataset2

Rule sequence	Rules	Class
R1	F49 <= 0	Low
R2	F49 > 0 & F1 >1 & F3 <= 1.44	Low
R3	F49 > 0 & F1 >1 & F3 > 1.44 & F37 <= 7	Low
R4 (H1)	F49 > 0 & F1 >1 & F3 > 1.44 & F37 > 7	High
R5 (H2)	F49 > 0 & F1 <= & F37 > 5	High
R6	F49 > 0 & F1 <= & F37 <=5 & F3 > 1.44	Low
R7	F49 > 0 & F1 <= & F37 <=5 & F3 <= 1.44 & F6 <=3	Low
R8 (H3)	F49 > 0 & F1 <= & F37 <=5 & F3 <= 1.44 & F6 > 3 & F6 <=5	High
R9	F49 > 0 & F1 <= & F37 <=5 & F3 <= 1.44 & F6 > 3 & F6 >5	Low

As shown in Figure 2, the starting point (first branch) of all decision tree rules is F49, which stands for prior online shopping experiences. As seen in Table 6, nine rules were extracted from the decision tree, of which three rules (R4, R5, and R8) are labeled as High. Thus, these are a significant group of factors that may predict heavy s-commerce use. In the following cases, the rules' effectiveness is estimated by the fulfillment of the factors given a specific number. The rules can be illustrated as follows:

— R4 (H1), being the first of the High class rules, it develops as follows: if F49 (representing prior experience in online shopping) is higher than 0 as a result of being represented by 1 (according to the key in Table 1, 1 means yes) and if F1 (representing gender) is higher than 1 (meaning it is 2 and according to the key in Table 1 means male) and F3 is higher than 1 (meaning it is 2 and according to Table 1 means single) and F37 is higher than 7 (thus, according to Table 1, the number of connections on SNSs exceeds 2 000), then the prediction of being a heavy social commerce user is high; three cases fulfill all the above factors.

In brief, R4 (H1) predicts that if users have prior online shopping experience and are single males who have more than 2 000 connections on SNSs, they are likely to be heavy SNS users; three cases reach the end of the branch. The same analysis is applied to all the rules whether the class is High or Low.

— The second High rule R5 (H2) predicts that users with prior online shopping experience who are females and have more than 500 SNS connections are likely to be heavy social commerce users, with an accuracy of 26:8 cases (out of 34 cases, 26 cases fulfill all the above factors and 8 do not).

— The last High-labeled rule R8 (H3) predicts that users with prior online shopping experience who are single females with a number of SNS connections equaling or less than 500 and an income between 300 and 1 000 JDs monthly are likely to be heavy social commerce users, with an accuracy of five cases fulfilling the rule.

— As shown previously, the Low class rules can be analyzed in the same manner.

The most important finding that we may extract from Figure 2 is that individuals with no prior experience in online shopping are less likely to be heavy social commerce users, with an accuracy of 180:37 cases (Rule 1).

5.2 Research Implications

The findings of this study have a number of important implications for future practice. In an immediate application, they can serve as guidelines for s-commerce community managers and firms planning to reap the benefits of investing in s-commerce. As the results show, factors such as gender, civil status, number of connections that users have on social networks, and prior experience in online shopping are among the most important components of s-commerce usage. These features can be rendered as base pillars for investors interested in s-commerce and its implementation in their businesses. As recommended by Chen and Shen (2015), businesses may consider new initiatives such as providing innovative one-click sharing functions on SNS platforms and encouraging the sharing of high-quality content through monetary or virtual rewards. Businesses may also use the features uncovered by this study to target s-commerce campaigns by using popular topics and expert contributors to reach out to potential consumers in a smart and innovative way. It is important to remember, however, that s-commerce facilitation is based on relational factors. Through the presented findings, companies could find new ways to develop and sustain mutually beneficial and long-term relation with s-commerce users.

This study has examined consumers' use of social commerce by analyzing different user features. A total of 49 features were extracted using data mining, which is a long-established approach for determining the significance of these features. A number of auto-selected features have been found to have a direct influence on s-commerce usage. Academic attention to these features, as well as investigations of other potentially influential features and their role in online shopping, will form a more comprehensive picture of s-commerce and further improve our current understanding of this emerging phenomenon. This research has uncovered many questions in need of further investigation. Typically, data mining aligns well with unstructured data. The data set analyzed in this research project was collected using an online survey, limiting the number of individual features that could be surveyed, collected, and then analyzed. More extensive work on the socio-demographic characteristics of online users and s-commerce participants is therefore required.

The findings of this research can be a milestone in shedding further light upon s-commerce data mining project implementations. Future research could include additional possibility changes in the data set, offering a new setting for the classification algorithms' parameters, and so on in an attempt to achieve highly accurate results and extract additional important information from the available data. To the best of our findings and knowledge, this implementation project is one of the pioneer studies using data mining techniques to examine data sets established based on an online survey to analyze human behavior in relation to specific technological use.

The key contribution of this study is paving the way for academics and researchers to use data mining in social science studies. It attempts to analyze the performance of classification algorithms used in data mining based on a data set collected via online survey to evaluate their utility in fulfilling the research project's goals and objectives. Further studies using this methodology should analyze the sufficiency of available data in producing reliable predictions, thus looking at any required changes during the data collection phase or any additional ways to improve the process.

Finally, more research and studies on the use of the data mining approach in social studies would help to establish a greater degree of accuracy in this realm. Combining the auto selection feature technique with a decision tree is strongly recommended as a means to explain feature selection results and, consequently, to evaluate users' behavior and technology adoption process. This type of research could help provide businesses' managers and academics with strategic and academic recommendations.

6. Conclusion

The proliferation of social networks has unleashed myriad new opportunities as well as new obstacles for researchers and businesspeople alike (Chen & Shen, 2015). For example, SNS-based market predictions are now more accurate due to information exchange in social networks (Qiu, Rui, & Whinston, 2013). This research project is an initial attempt to explore individuals' social commerce usage intention using a data mining approach. This project's findings were obtained by applying data mining classification algorithms and an auto feature selection method to the collected data because each classifier of the five under consideration proceeds differently. We have noticed that the data related to gender, civil status, monthly income, number of connections, and prior online shopping experience are among the major factors influencing the classification process.

In the applied realm, these findings will assist practitioners in improving their future social commerce strategies. Meanwhile, these findings can also guide researchers' investigations of social commerce issues. We believe this project to be a primary step toward better applying data mining in social science studies because the methodology employed can be further used in laying the first brick for additional social science data mining implementation projects.

Acknowledgments

References

Aha, D. W., Kibler, D., & Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learning*, *6*(1), 37-66. https://doi.org/10.10 07/BF00153759

Al-Zoubi, A., Alqatawna, J., & Faris, H. (2017). Spam Profile Detection in Social Networks Based on Public Features. *International Conference on Information and Communication Systems* (pp. 130-135). Amman: IEEE. Retrieved from https://www.researchgate.net/profile/Ala_Al-Zoubi/publication/316973828_Spam_Profile_Detection_in_So cial_Networks_Based_on_Public_Features/links/591b0be4458515bb03cde5a1/Spam-Profile-Detection-in-S ocial-Networks-Based-on-Public-Features.pdf

Awad, N. F., & Ragowsky, A. (2008). Establishing trust in electronic commerce through online word of mouth: An examination across genders. *Journal of Management Information Systems*, *24*(4), 101-121. https://doi.org/10.2753/MIS0742-1222240404

Barclays. (2012). Social commerce: The next generation of retail. Available at http:// www.barclayscorporate.com/content/dam/corppublic/corporate/Documents/re search/S-Commerce-report.pdf

Barnes, N. G. (2014). Social commerce emerges as big brands position themselves to turn" follows"," likes" and" pins" into sales. *American Journal of Management*, *14*(4), 11.

Brunk, C. A., & Pazzani, M. J. (1991). An investigation of noise-tolerant relational concept learning algorithms. In *Machine Learning Proceedings 1991* (pp. 389-393). https://doi.org/10.1016/B978-1-55860-200-7.50080-5

Campbell, C. (2000). Shopaholics, spendaholics, and the question of gender. *I shop, therefore I am: Compulsive buying and the search for self*, 57-75.

Chang, Y. T., Yu, H., & Lu, H. P. (2015). Persuasive messages, popularity cohesion, and message diffusion in social media marketing. *Journal of Business Research*, *68*(4), 777-782.

Chaykowski, K. (2016, September 27). *Sheryl Sandberg: Facebook's 4 Million Advertisers Are 'Proof' Of the Power of Mobile.* Retrieved from https://www.forbes.com/sites/kathleenchaykowski/2016/09/27/sheryl-sandberg-facebooks-4-million-adverti sers-are-proof-of-the-power-of-mobile/#29e491ab1f17

Chaykowski, K. (2015, December 8). *Number of Facebook Business Pages Climbs To 50 Million With New Messaging Tools.* Retrieved from https://www.forbes.com/sites/kathleenchaykowski/2015/12/08/facebook-business-pages-climb-to-50-millio n-with-new-messaging-tools/#2ac7855d6991

Chen, A., Lu, Y., & Wang, B. (2017). Customers' purchase decision-making process in social commerce: A social learning perspective. *International Journal of Information Management*, *37*(6), 627-638.

Chen, J., & Shen, X. L. (2015). Consumers' decisions in social commerce context: An empirical investigation. *Decision Support Systems*, *79*, 55-64. https://doi.org/10.1016/j.dss.2015.07.012

Chiu, C., I, C., & Silverman, A. (2012). Understanding social media in China. *McKinsey Quarterly*, *2*(2012), 78-81.

Curty, R. G., & Zhang, P. (2011). Social commerce: Looking back and forward. *Proceedings of the American Society for Information Science and Technology*, *48*(1), 1-10.

Dittmar, H., Long, K., & Meek, R. (2004). Buying on the Internet: Gender differences in on-line and conventional buying motivations. *Sex roles*, *50*(5-6), 423-444.https://doi.org/10.1023/B:SERS.0000018896.35251.c7

Faris, H., Ala'M, A.Z. and Aljarah, I., 2017, October. Improving email spam detection using content based feature engineering approach. In *Applied Electrical Engineering and Computing Technologies (AEECT), 2017 IEEE Jordan Conference on* (pp. 1-6). IEEE.

Faris, H., Mafarja, M. M., Heidari, A. A., Aljarah, I., Ala'M, A. Z., Mirjalili, S., & Fujita, H. (2018). An Efficient Binary Salp Swarm Algorithm with Crossover Scheme for Feature Selection Problems. *Knowledge-Based Systems*.

Fürnkranz, J., & Widmer, G. (1994). Incremental reduced error pruning. In *Machine Learning Proceedings 1994* (pp. 70-77). https://doi.org/10.1016/B978-1-55860-335-6.50017-9

Gefen, D., & Straub, D. W. (2000). The relative importance of perceived ease of use in IS adoption: A study of e-commerce adoption. *Journal of the association for Information Systems*, *1*(1), 8.

Hajli, M. (2013). A research framework for social commerce adoption. *Information Management & Computer Security*, *21*(3), 144-154. https://doi.org/10.1108/IMCS-04-2012-0024

Hajli, M. N. (2014). The role of social support on relationship quality and social commerce. *Technological Forecasting and Social Change, 87,* 17-27. https://doi.org/10.1016/j.techfore.2014.05.012

Hajli, N. (2015). Social commerce constructs and consumer's intention to buy. *International Journal of Information Management*, *35*(2), 183-191.https://doi.org/10.1016/j.ijinfomgt.2014.12.005

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

Hennig-Thurau, T., Hofacker, C. F., & Bloching, B. (2013). Marketing the pinball way: understanding how social media change the generation of value for consumers and companies. *Journal of Interactive Marketing*, *27*(4), 237-241.

Huang, Z., & Benyoucef, M. (2013). From e-commerce to social commerce: A close look at design features. *Electronic Commerce Research and Applications*, *12*(4), 246-259.

Immediate Future (2010). *The explosion of social shopping. London*. Retrieved from http://immediatefuture.co.uk/

Insights, H. (2014). Social commerce strategy and outlook. *2014-01-30)[2016-03-04]*. Retrieved from http://www.prlog.org/11872607-hnyb-insights-predicts-global-social-commerce-market-totouch-80bn-by-2020. Html

Jascanu, N., Jascanu, V., & Nicolau, F. (2007). A new approach to E-commerce multi-agent systems.

Kabakchieva, D. (2013). Predicting student performance by using data mining methods for classification. *Cybernetics and information technologies*, *13*(1), 61-72.

Kemp, S. (2018). Digital in 2018: world's Internet users pass the 4 billion mark. *New York, We Are Social*, *30*.

Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S. (2011). Social media? Get serious! Understanding the functional building blocks of social media. *Business horizons*, *54*(3), 241-251.

Kim, S., & Park, H. (2013). Effects of various characteristics of social commerce (s-commerce) on consumers' trust and trust performance. *International Journal of Information Management*, *33*(2), 318-332.

Li, Z., & Li, C. (2014). Twitter as a social actor: How consumers evaluate brands differently on Twitter based on relationship norms. *Computers in Human Behavior*, *39*, 187-196.

Liébana-Cabanillas, F., & Alonso-Dos-Santos, M. (2017). Factors that determine the adoption of Facebook commerce: The moderating effect of age. *Journal of Engineering and Technology Management*, *44*, 1-18.

Liu, H., Chu, H., Huang, Q., & Chen, X. (2016). Enhancing the flow experience of consumers in China through

interpersonal interaction in social commerce. *Computers in Human Behavior*, *58*, 306-314.

Mathison, R. (2018, January 24). *23+ Useful Instagram Statistics for Social Media Marketers.* Retrieved from https://blog.hootsuite.com/instagram-statistics/

McCarthy, J. (2015, 24 April). *Facebook influences over half of shoppers says DigitasLBi's Connected Commerce report.* Retrieved from http://www.thedrum.com/news/2015/04/24/facebook-influences-over-half-shoppers-says-digitaslbi-s-connected-commerce-report.

Pagallo, G., & Haussler, D. (1990). Boolean feature discovery in empirical learning. *Machine learning*, *5*(1), 71-99. https://doi.org/10.1007/BF00115895

Paola, J. D., & Schowengerdt, R. A. (1995). A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification. *IEEE Transactions on Geoscience and Remote sensing*, *33*(4), 981-996.

Poushter, J. (2016). Smartphone ownership and internet usage continues to climb in emerging economies. *Pew Research Center*, *22*.

Qiu, L., Rui, H., & Whinston, A. (2013). Social network-embedded prediction markets: The effects of information acquisition and communication on predictions. *Decision Support Systems*, *55*(4), 978-987.https://doi.org/10.1016/j.dss.2013.01.007

Quinlan, J. R. (1987). Simplifying decision trees. *International journal of man-machine studies*, *27*(3), 221-234. https://doi.org/10.1016/S0020-7373(87)80053-6

Quinlan, J. R. (1993). Constructing decision tree. *C4*, *5*, 17-26.

Rapp, A., Beitelspacher, L. S., Grewal, D., & Hughes, D. E. (2013). Understanding social media effects across seller, retailer, and consumer interactions. *Journal of the Academy of Marketing Science*, *41*(5), 547-566.

Rodgers, S., & Harris, M. A. (2003). Gender and e-commerce: An exploratory study. *Journal of advertising research*, *43*(3), 322-329.

Safavian, S. R., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics*, *21*(3), 660-674.

Sharma, S., & Crossler, R. E. (2014). Intention to engage in social commerce: uses and gratifications approach.

Sturiale, L., & Scuderi, A. (2013). Evaluation of social media actions for the agrifood system. *Procedia Technology*, *8*, 200-208. https://doi.org/10.1016/j.protcy.2013.11.028

Wigand, R. T., Benjamin, R. I., & Birkland, J. L. (2008, August). Web 2.0 and beyond: implications for electronic commerce. In *Proceedings of the 10th international conference on Electronic commerce* (p. 7). ACM.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Wu, Y. C. J., Shen, J. P., & Chang, C. L. (2015). Electronic service quality of Facebook social commerce and collaborative learning. Computers in Human Behavior, 1-8. https://doi.org/10.1016/j.chb.2014.10.001

Xiang, L., Zheng, X., Lee, M. K., & Zhao, D. (2016). Exploring consumers' impulse buying behavior on social commerce platform: The role of parasocial interaction. *International Journal of Information Management*, *36*(3), 333-347.

Zare, M., Pourghasemi, H. R., Vafakhah, M., & Pradhan, B. (2013). Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network model: a comparison between multilayer perceptron (MLP) and radial basic function (RBF) algorithms. *Arabian Journal of Geosciences*, *6*(8), 2873-2888. https://doi.org/10.1007/s12517-012-0610-x

Zeng, F., Huang, L., & Dou, W. (2009). Social factors in user perceptions and responses to advertising in online social networking communities. *Journal of interactive advertising*, *10*(1), 1-13.

Zhang, H., Lu, Y., Gupta, S., & Zhao, L. (2014). What motivates customers to participate in social commerce? The impact of technological environments and virtual customer experiences. *Information & Management*, *51*(8), 1017-1030.

Zhang, K. Z., & Benyoucef, M. (2016). Consumer behavior in social commerce: A literature review. *Decision*

Support Systems, *86*, 95-108.

Zhang, K. Z., Benyoucef, M., & Zhao, S. J. (2015). Consumer participation and gender differences on companies' microblogs: A brand attachment process perspective. *Computers in Human Behavior*, *44*, 357-368.

Zheng, X., Zhu, S., & Lin, Z. (2013). Capturing the essence of word-of-mouth for social commerce: Assessing the quality of online e-commerce reviews by a semi-supervised approach. *Decision Support Systems*, *56*, 211-222.

Copyrights