Predictions on Usefulness and Popularity of Online Reviews: Evidence from Mobile Phones for Older Adults

Minghuan Shou(✉), Xueqi Bao, and Jie Yu

Business School, University of Nottingham Ningbo China, Ningbo 315100, China
minghuan.shou@nottingham.edu.cn

Abstract. This paper aims to propose an effective method to locate valuable reviews of mobile phones for older adults. After collecting the online reviews of mobile phones for older adults from JD mall, we propose a three-step framework. Firstly, Topic Modeling models and linguistic inquiry and word count (LIWC) methods are employed to extract latent topics. Secondly, regression models are used to examine the effect of variables obtained from the first step on the popularity (number of replies) and usefulness (number of helpful counts). Thirdly, seven machine learning models are adopted to predict the popularity and usefulness of online reviews. The results indicate that although older adults are more interested in the exterior, sound, money, and communication functions of mobile phones, they still care about the touch feel, work, and leisure functions. In addition, Random Forest performs the best in predicting the popularity and usefulness of online reviews. The findings can help e-commerce platforms and merchants identify the needs of the targeted consumers, predict which reviews will get more attention, and provide some early responses to some questions.

Keywords: Online review · Older adult · Mobile phone

1 Introduction

Information and communication technologies (ICT) have shifted the social interactions (Jara et al. 2015) and the fabric of economic life (Song et al. 2020). However, serious inequalities in access to and use of ICT have appeared among the categories of persons in one group (Park et al. 2015). For example, some of the public may have problems making use of the various digital devices. Scholars call this unequal disparity in ICT diffusion a digital divide. Whereas the digital divide occurring in society can create and aggravate economic as well as social inequalities. It has also been regarded as one of the most important barriers to social inclusion and fostering a strong and creative economy (Park et al. 2015). Thus, it is urgent to propose some methods to solve the digital divide. Digital divides are now understood to be complicated with several dimensional situations (Cruz et al. 2017) like accessibility, affordability, reliability, speed, and utilization (Loo 2012). More seriously, digital divides have been proved to exist in many countries (Szeles 2018) and have become a global problem. Current findings of the influential factors mainly

M. Kurosu et al. (Eds.): HCII 2022, LNCS 13516, pp. 475–489, 2022.
https://doi.org/10.1007/978-3-031-17615-9_33

contain gender (Mumporeze and Prieler 2017; Potnis 2016), age (Ball et al. 2019; Hall et al. 2015), economic development (Fuchs 2009), and education (Estacio et al. 2019).

On the other hand, the increasing number of older adults (aged 65+) makes the age-based digital divide more serious. Older adults always discontinue adopting digital communication technologies (Neves et al. 2018). For example, older people in Greece and Bulgaria are more than 11 times less likely to be online than the overall population (Niehaves and Plattfaut 2014). Additionally, a survey exploring computer and internet experience among younger adults (n = 430, aged 18–28 years) and older adults (n = 251, aged 65–90 years) found the difference significant. The percentages of older people showing experience with computers and the internet are 80% and 50%, respectively, compared to 99% and 90% for the younger adults (Olson et al. 2011). Psychological aging theory and processing-speed theory suggest that older adults may have some cognitive limitations on using new technologies, which explains the lower proportion of older adults with computer and internet experience.

Since only very few older adults are using the internet and the number of older adults is rapidly growing, mobile phone designers and merchants can make more profits if the future products can attract older adults. Besides, online reviews are regarded as a source of information for decision-making because of the abundance and ready availability of information. However, the sheer volume of online reviews makes it hard for consumers, especially older adults who perceive more difficulties in reading reviews and obtaining information compared to younger adults, to locate the useful ones. If merchants can locate the online reviews, which will receive more attention from the public in advance, they can increase older adults' satisfaction by placing the useful ones in a prominent place to help the public get better access to information and save time.

Thus, there are two main purposes of this paper. The first aim is to understand what characteristics of the technology are applicable to older adults, and the second one is to predict the popularity (number of replies) and usefulness (number of helpful counts) of online reviews via machine learning models. A three-step research framework is designed. Firstly, we crawl the online reviews of mobile phones from JD Mall (https://www.jd.com/), and then construct the Topic Modeling models and linguistic inquiry and word count (LIWC) to investigate what characteristics older adults prefer. Secondly, regression models are employed to explore what characteristics obtained from the first step can significantly affect the popularity (number of replies) and usefulness (number of helpful counts) of online reviews. Thirdly, machine learning models are adopted to predict the popularity and usefulness of online reviews based on their significantly influential characteristics.

Our study deviates from previous studies in some aspects. The main contributions are as follows.

Firstly, this paper provides a guideline for scholars on how to solve the age-digital divide. By adopting Topic Modelling models and analyzing the data of online reviews, this paper accurately obtains the characteristics of mobile phones older adults prefer. Scholars can take advantage of the preference to continue their studies and avoid some useless experiments. This can effectively save time and resources. Besides, because the age-digital divide has not currently been solved, merchants struggle to explore what product value customers want. What this paper finds can provide evidence for designers by clearly showing the contribution of comment information to the public attention, and merchants can make use of the results to design the products.

Secondly, the findings of this paper argue that the cognitive limitations of older adults on new technologies are decreasing. This provides an oppositive view to the current theories which present that older adults are more likely to refuse to use new technologies. Besides, the findings can also provide support to the work published by Ghasemaghaei et al. (2019) that older adults do not always mean simpler IT products. In some cases, they may perceive IT products with other features.

Thirdly, the results of this paper indicate that it can be possible to locate useful online reviews in advance via machine learning models. For merchants, they can better manage their time and effort in dealing with consumers' responses based on the results of predictive models. In addition, they can be alerted and take precautions measures against negative consumer comments.

The paper is structured as follows: the following section shows the theoretical foundation of this paper and some studies on how to solve the digital divide. Section 3 details the research framework and the data information. A presentation of results follows. The conclusion and some suggestions are drawn in Sect. 5. Finally, this paper summarizes the limitations and the future work.

2 Literature Review

This section covers two parts. The first part is a summary of the previous studies on the digital divide. The second part is to present the development of the model we use in this article.

2.1 Studies About the Digital Divide

Just as we can see from section one, the digital divide has become a global problem, and its development of it has strongly affected social and economic development. Thus, it is urgent for scholars to think of some methods to change the status. However, by summarizing the previous studies, we are surprised to find that there are only a few studies on how to solve the problem, and the existing papers mainly focus on the following two aspects.

On the one hand, scholars spent much time building models to evaluate the digital divide. Pertrovic et al. (2012) proposed an innovative procedure for benchmarking the digital divide. Classifying countries into hierarchical levels of their performance and selecting the benchmarks for less successful ones are two main innovations. Cruz-Jesus et al. (2012) addressed the European digital disparities by building multivariate statistical models and cluster analysis. Park et al. (2015) and Billon et al. (2010) also dedicated to studying the digital divide between countries. But they took advantage of the similar method used by Pertrovic et al. (2012) as well to categorize the countries into several different groups and explained the specific appearance of the digital divide in each group. Kiss et al. (2020) and Chang et al. (2012) both collected data by questionnaire, but the first one took advantage of the cluster method, and the latter chose to use the analytic hierarchy process (AHP).

On the other hand, exploring the conditions where the digital divide may occur is also one of the focuses. Except for the factors mentioned in section one, there are more others. Reddick et al. (2020) conducted a survey to explore the influence of affordability factors on the broadband digital divide. They found that geographical disparities play an important role. Besides, the digital divide was not exclusively a rural/urban digital divide but also appeared in an intra-city context. Jiang et al. (2019) paid attention to analyzing the appearance of cancer survivors. They found that some patients had problems using online patient-provider communication (OPPC), which meant that there were significant digital divide barriers existing in the adoption and actual usage of OPPC. Duplaga (2017) and Sachdeva et al. (2015) also found evidence of a digital divide among patients.

By reading the papers about the solutions to the digital divide, we find that the existing research mainly focused on providing general suggestions for the governments. Akca et al. (2007) and Huang and Cox (2014) proposed the design of web pages and internet access could be a method. Loo and Ngan (2012) suggested that governments should implement some measures to encourage the use of mobile phones and personal computers. Similarly, Townsend et al. (2013) also mentioned that a growing digital divide would make it more dangerous, and increasing broadband usage would be necessary.

As a basis for summarizing the above papers, it was not difficult to find that there were so many methods to evaluate the digital divide and the digital divide occurred in various conditions. However, only a few scholars focused on providing evidence for governments to make policies to solve the digital divide, but their suggestions were not targeted. Thus, it was urgent to explore some new methods.

2.2 Researches on Online Reviews

Online reviews have become indispensable information sources that can provide useful information to customers (Mudambi and Schuff 2010). Customers can obtain information about the products that they want to buy before purchasing from either the review content or the communicative interaction between the product review author and other consumers (Cao 2020; Li and Huang 2020). Due to the sufficient information from the online reviews, scholars have started to build some models to make some analyses on them, aiming at obtaining the preference of customers. Wang et al. (2020) developed a novel approach based on Bayesian statistics to explore the prior belief in online reviews and solved the problem that review-hosting firms could not detect the helpfulness of reviews. Choi et al. (2020) created an aging theory-based EDT algorithm to build the groups of similar reviews and tracked their development. Finally, by using sentiment analysis and evaluating time-evolving events, people could identify time-evolving product opportunities. Linguistic issues, sentiment analysis, and content analysis were also performed by Teso et al. (2018). Besides, they suggested that retailers could adapt the offer to the gender of customers. In summary, online reviews obtained much information, and taking advantage of the information could bring more opportunities.

3 Research Framework

This section introduces our research model and methods for data collection, processing, and measurement (Fig. 1). Before making some analyses, we plan to mine the online reviews of mobile phones for older adults from JD mall. LDA and NMF methods can be performed to reveal the latent dimensions that express the customer's priorities. These dimensions present what customers prefer while purchasing the products. Besides, the results of LIWC can help us test the results obtained by the Topic Modelling models. After that, this paper explores whether the characteristics obtained from the Topic Modeling models can significantly affect the popularity and usefulness of online reviews. Eventually, seven machine learning models are employed to predict the popularity and usefulness of online reviews. To display the results, we divide the entire data into two parts based on the median of the output features. For example, the median output (number of replies to online reviews) is 5, so this paper converts the number (>= 5) to 1 and the number (<5) to 0, representing 43.67% and 56.33%, respectively. The median output (number of helpfulness) is 4, so this paper converts the number (>= 4) to 1 and the number (<4) to 0, representing 38.19% and 61.81%, respectively.

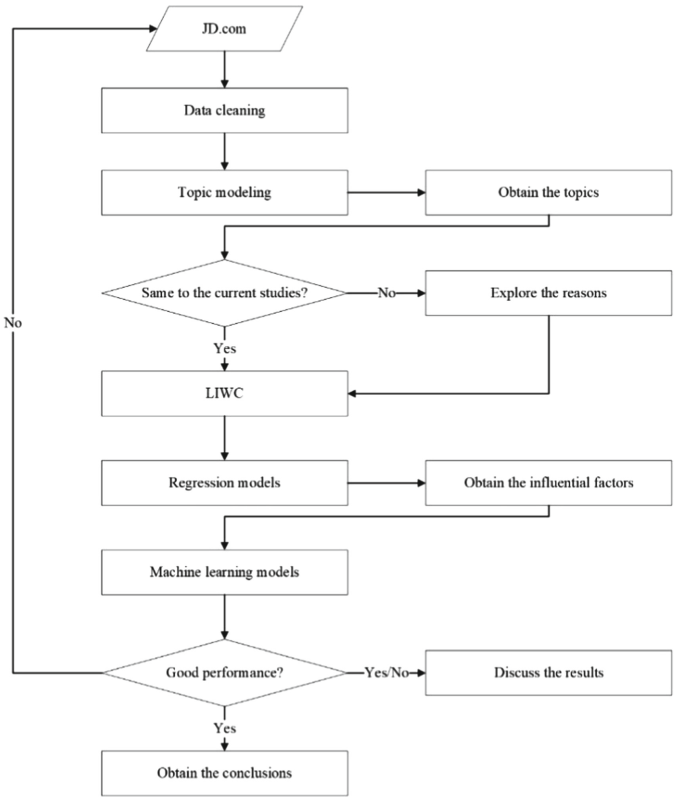

Fig. 1. Research framework

A web crawler has been designed to obtain the online reviews of mobile phones for elders from JingDong mall. Some data cleaning methods have been performed. We mainly followed the steps suggested in (Tirunillai and Tellis 2014). In the first step, some irrelevant characters and words, such as HTML, tags, URLs, and punctuation marks are eliminated due to their no information content about the product or the dimensions of quality that we are interested in extracting. After that, the presence of characters (e. g. ".", "!", "?") is used to break the reviews into several sentences. Then we apply the part-of-speech method to retain words that are adjectives, nouns, or adverbs. The above steps are to collect the words with sufficient information about the product and its quality. Finally, stop-words are also deleted, and function words are numbered to support the following analysis. A total of 284527 comments have been collected after data cleaning.

Besides, seven popular machine learning models were chosen to predict the "usefulness" and "popularity" of online reviews, including Random Forest, Logistic Regression, Decision Tree, Naive Bayes, Neural Network, Support Vector Machine (SVM), and K-Nearest Neighbor (KNN). Their performances were mainly measured by accuracy. The confusion matrix is shown in Table 1. Besides, we also calculate AUC and draw ROC curves to compare the results.

Table 1. Confusion matrix

		Prediction value	
		Positive	Negative
Real value	Positive	True positive (TP)	False negative (FN)
	Negative	False positive (FP)	True negative (TN)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

4 Results

4.1 Results of Topic Modeling Models

Two different models (LDA & NMF) are employed to obtain what characteristics of mobile phones older adults prefer, which can prove the robustness of the results. To foster our understanding, we also divide the whole data into two groups based on the number of stars. We only show the first nine topics summarized from the detailed information to show the results clearly.

Table 2. Results of topic modeling model (LDA & NMF)

	Data		Data (Star $>\ = 3$)		Data (Star < 3)	
Topic	NMF	LDA	NMF	LDA	NMF	LDA
1	Function	Battery	Function	Old	Function	Quality

(*continued*)

Table 2. (*continued*)

	Data		Data (Star > = 3)		Data (Star < 3)	
Topic	NMF	LDA	NMF	LDA	NMF	LDA
2	Exterior	Price	Operation	Battery	Operation	Sound
3	Feel	Communication	Feel	Communication	Feel	Communication
4	Money	Old	Battery	Money	Exterior	Operation
5	Sound	Exterior	Sound	Exterior	Sound	Storage
6	Delivery	Delivery	Parent	Sound	Delivery	Delivery
7	Parent	Feel	Delivery	Signal	Parent	Function
8	Operation	Parent	Exterior	Delivery	Font	Exterior
9	Quality	Sound	Money	Function	Quality	Price

Table 2 shows the detailed information obtained from two topic modeling models (LDA & NMF). There is nearly no difference between the results from the two models, which means that it is reasonable to make some analysis on these topics. The results indicate that older adults are still concerned more about the communication function, exterior, touch feel, and sound of mobile phones. Besides, they are still concerned about the work, leisure, and money of the products.

This paper considers using LIWC to get consumers' attention from online reviews. The results of this component can help us verify the conclusions obtained above. The findings are shown in Fig. 2 support the results of Topic modeling. While among these seven features, older adults pay more attention to the money, sound, exterior, and communication function than the rest.

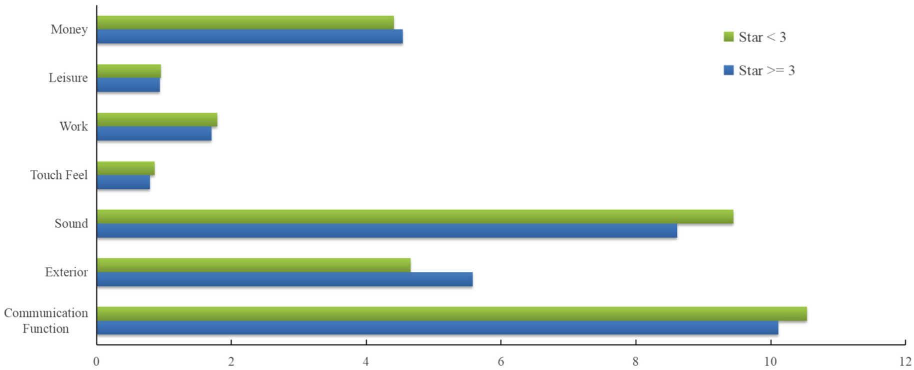

Fig. 2. Results of LIWC

4.2 Results of Regression Models

Before employing these features obtained from Topic Modeling models to predict the popularity and usefulness of online reviews, regression models are adopted to examine whether these features have a significant effect.

As shown in Table 3, the average word count of the comment is 14.53, and the variables (7–13) refer to consumers' attention towards each aspect. According to the correlation analysis results, there is no strong relationship between the independent variables, which means the regression results are not affected by the multicollinearity when predicting the popularity and usefulness of online reviews.

Table 3. Statistical description and correlation analysis of the variables

		Mean	1	2	3	5	6	7	8	9	10	11	12	13
1	Star	3.52	1.00											
2	Popularity	0.44	−0.77	1.00										
3	Usefulness	0.38	−0.16	−0.07	1.00									
5	Sentiment	0.79	0.11	0.04	−0.05	1.00								
6	Word Count	14.53	0.00	−0.10	−0.06	0.10	1.00							
7	Communication Function	10.33	−0.03	0.04	0.01	−0.01	−0.03	1.00						
8	Exterior	5.1	0.06	−0.04	−0.02	0.22	0.04	−0.12	1.00					
9	Sound	9.04	−0.03	0.05	0.00	−0.01	−0.01	0.46	0.04	1.00				
10	Touch Feel	0.83	−0.02	0.00	0.01	−0.03	−0.01	−0.03	−0.03	−0.05	1.00			
11	Work	1.74	−0.01	0.01	−0.01	0.01	0.03	0.01	−0.07	−0.05	−0.01	1.00		
12	Leisure	0.95	−0.01	0.00	0.01	0.01	0.01	0.04	−0.05	0.00	0.00	0.01	1.00	
13	Money	4.47	0.00	0.02	0.00	0.05	−0.04	0.08	−0.13	−0.10	−0.03	0.04	0.07	1.00

In the regression models, three control variables are located, including star, sentiment, and word count. After examining the effect of control variables on the dependent ones (column (1) in Tables 4 and 5), the effect of the seven independent variables is tested (columns (2–8) in Tables 4 and 5). Before examining whether these variables can significantly affect the popularity and usefulness of online reviews, the significance of the models is tested first. In Tables 4 and 5, all the models are statistically significant at a 0.01 level, and all the control variables are significantly related to the dependent ones. When we predict the popularity of online reviews, we do find that only six independent variables show a significant effect, including the exterior, sound, touch feel, leisure, money, and communication functions. When we predict the usefulness of online reviews, only sound, touch feel, work, and leisure show a significant effect. Accordingly, we combine all the control variables and the significant independent variables to construct machine learning models.

Table 4. Results of predicting the popularity (number of replies) of online reviews

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Exterior		−0.0010***						
		(0.0001)						
Sound			0.0013***					
			(0.0001)					
Touch feel				−0.0008***				
				(0.0001)				
Work					0.0001			
					(0.0001)			
Leisure						−0.0010***		
						(0.0001)		
Money							0.0005***	
							(0.0001)	
Communication function								0.0007***
								(0.0000)
Star	−0.2445***	−0.2442***	−0.2242***	−0.2445***	−0.2445***	−0.2445***	−0.2445***	−0.2243***
	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Sentiment	0.2190***	0.2253***	0.2191***	0.2188***	0.2190***	0.2192***	0.2183***	0.2190***
	(0.0018)	(0.0019)	(0.0018)	(0.0018)	(0.0018)	(0.0018)	(0.0018)	(0.0018)
Word count	−0.0016***	−0.0016***	−0.0016***	−0.0016***	−0.0016***	−0.0016***	−0.0016***	−0.0016***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
_cons	1.1475***	1.1469***	1.1344***	1.1485***	1.1473***	1.1484***	1.1455***	1.1399***
	(0.0019)	(0.0019)	(0.0020)	(0.0019)	(0.0019)	(0.0019)	(0.0019)	(0.0020)
R-square	0.6201	0.6204	0.621	0.6201	0.6201	0.6201	0.6202	0.6204
Prob > F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Number of obs	284527	284527	284527	284527	284527	284527	284527	284527

Notes: N = 284527. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.3 Results of Machine Learning Models

Seven machine learning models, namely Random Forest, Logistic Regression, Decision Tree, Naïve Bayes, Neural Network, SVM, and KNN, are employed in this part. Table 6 and Figs. 3 and 4 present that the accuracy values of all models are higher than 0.9, suggesting that machine learning models can be well used to predict the popularity of online reviews based on the number of stars, sentiment, and word counts of online reviews. In addition, after combining the control variables with independent variables, the performance of the majority of the machine learning models becomes better, except for the SVM and KNN when predicting the popularity of online reviews.

Table 5. Results of predicting the usefulness (number of helpful counts) of online reviews

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Exterior		0.0001						
		(0.0001)						
Sound			−0.0003***					
			(0.0001)					
Touch feel				0.0007***				
				(0.0002)				
Work					−0.0011***			
					(0.0002)			
Leisure						0.0006**		
						(0.0002)		
Money							−0.0000415	
							(0.0001)	
Communication function								5.73E-06
								(0.0001)
Star	−0.0498***	‘−0.0498***	− 0.0499	−0.0498***	−0.0499***	−0.0498***	−0.0498***	−0.0498***
	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Sentiment	−0.0400***	’−0.0406***	−0.0400***	−0.0398***	−0.0398***	−0.0401***	−0.0399***	−0.0400***
	(0.0029)	(0.0029)	(0.0029)	(0.0029)	(0.0029)	(0.0029)	(0.0029)	(0.0029)
Word count	−0.0009***	−0.0009***	−0.0009***	−0.0009***	−0.0009***	−0.0009***	−0.0009***	−0.0009***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
_cons	0.6010***	0.6011***	0.6038***	0.6002***	0.6029***	0.6005***	0.6011***	0.6009***
	(0.0030)	(0.0030)	(0.0031)	(0.0030)	(0.0030)	(0.0030)	(0.0030)	(0.0031)
R-square	0.0324	0.0324	0.0324	0.0324	0.0325	0.0324	0.0324	0.0324
Prob > F	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Number of obs	284527	284527	284527	284527	284527	284527	284527	284527

Notes: N = 284527. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6. Results of machine learning models on predicting the popularity of online reviews

	Control variables		Control variables + independent variables		Difference	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
Random forest	96.11%	96.07%	96.95%	96.94%	0.84%	0.87%
Logistic regression	92.83%	93.24%	92.94%	93.30%	0.11%	0.06%
Decision tree	96.06%	96.00%	96.39%	96.33%	0.33%	0.33%
Naive bayes	93.43%	94.03%	93.43%	94.03%	0.00%	0.00%
Neural network	94.10%	94.18%	94.40%	94.61%	0.30%	0.43%
SVM	92.48%	93.13%	92.40%	93.09%	−0.08%	−0.04%
KNN	94.58%	94.57%	90.76%	90.68%	−3.82%	−3.89%

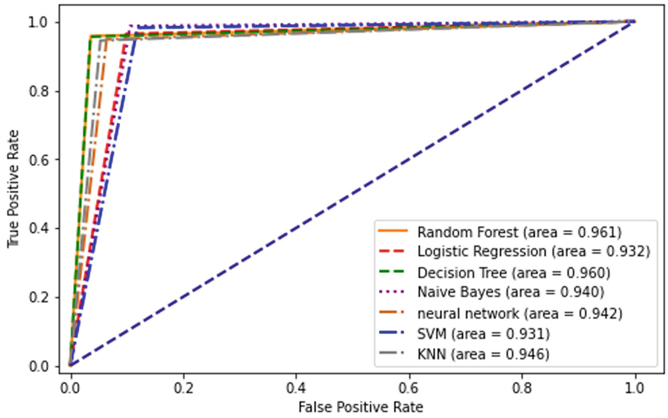

Fig. 3. Results of predicting the popularity of online reviews with control variables

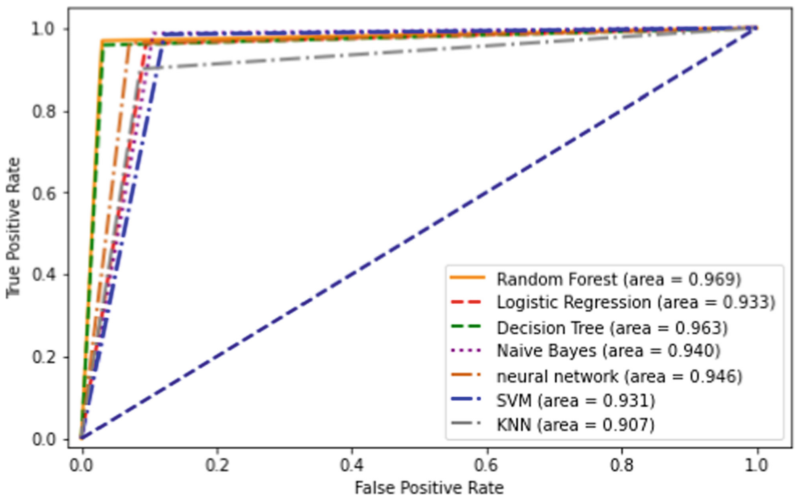

Fig. 4. Results of predicting the popularity of online reviews with control variables and independent variables

Table 7 and Figs. 5 and 6 focus on predicting the usefulness of online reviews via seven machine learning models. The results indicate that there is a large difference in the performance of machine learning models, with the Random Forest and Decision Tree performing the best. In addition, after combining the control variables with the independent variables to construct the machine learning models, we do find that for Random Forest, Logistic Regression, Decision Tree, Naïve Bayes, Neural Network, and KNN, the improvement is significant and positive. While for SVM, the effect is diverse.

Table 7. Results of machine learning models on predicting the usefulness of online reviews

	Control variables		Control variables + independent variables		Difference	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
Random forest	77.01%	74.88%	78.56%	77.02%	1.55%	2.14%
Logistic regression	62.00%	50.14%	62.03%	50.25%	0.03%	0.11%
Decision tree	77.09%	75.09%	78.36%	76.91%	1.27%	1.82%
Naive bayes	51.20%	53.93%	53.37%	55.29%	2.17%	1.36%
Neural network	63.26%	53.45%	63.42%	54.44%	0.16%	0.99%
SVM	51.43%	50.09%	44.31%	49.53%	–7.12%	–0.56%
KNN	68.49%	66.12%	68.52%	66.25%	0.03%	0.13%

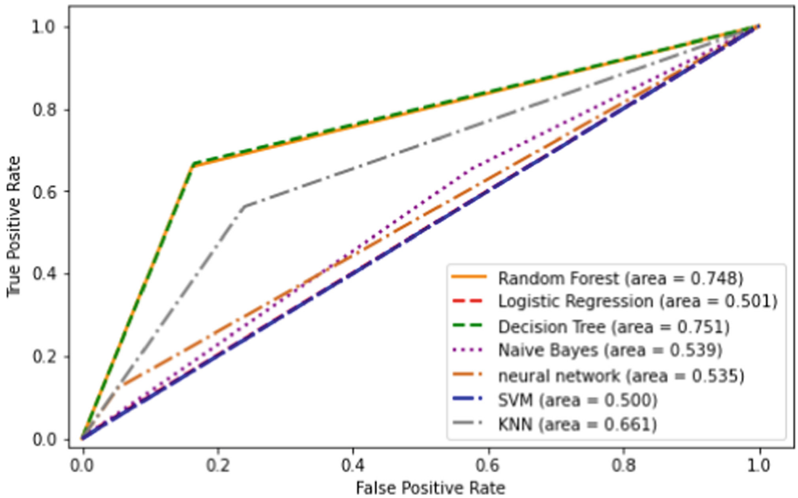

Fig. 5. Results of predicting the usefulness of online reviews with control variables

5 Discussions and Limitations

After conducting the Topic Modeling models, we obtain that the older adults not only care about the exterior, communication function, touch feel, sound, and money of the mobile phones but also pay attention to the leisure and work. Although the current theories state that older adults have cognitive limitations on using new technologies, what this paper obtains proposes an opposite view. Older adults' requirement for the multi-functions of mobile phones is rising. Traditional mobile phones, which can only be used for phone calls, can no longer meet the requirements of older adults nowadays. The findings also provide support to the work published by Ghasemaghaei et al. (2019). Two possible reasons can be used to explain it. Firstly, economic development provides

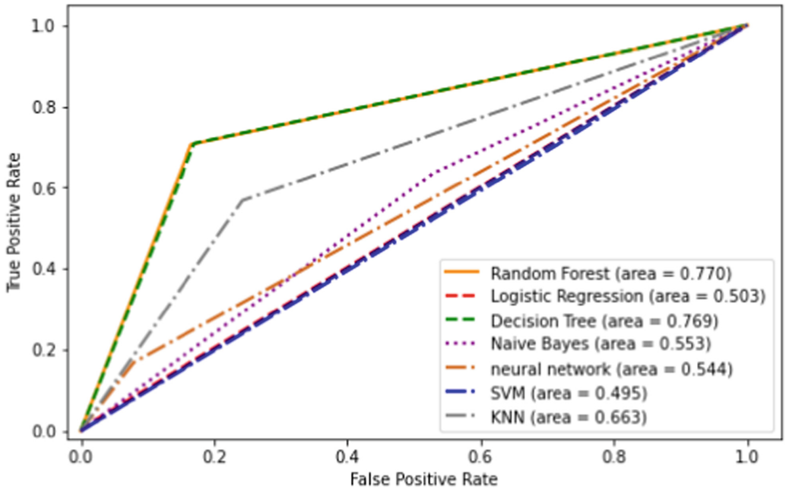

Fig. 6. Results of predicting the usefulness of online reviews with control variables and independent variables

older adults more opportunities to access new technologies, which can contribute to the reduction in their cognitive limitations. Secondly, public library constructions can provide rich physical infrastructural resources and free training sessions for the public, which can help reduce the digital divide (Manzuch and Maceviciute 2020). Older adults have more free time to access the resources provided by libraries than younger ones, suggesting that the age-based digital divide can be well reduced.

The main limitation of this paper is that the data collection in this paper is mainly on smart mobile phones. Due to the lack of data about non-smart mobile phones, some comparisons cannot be made. This means that this study cannot investigate whether there is a significant difference in product value between smart mobile phones and non-smart ones. The difference can be meaningful for increasing the sales of smart mobile phones. To solve this problem, collecting more data is necessary. Making some comparisons between phones with different product values can be beneficial for obtaining the effect of the product value. Besides, designing some experiments to explore the influence mechanism of these two variables while people purchase mobile phones can be useful.

6 Conclusion and Suggestions

The findings indicate that currently, although older adults pay more attention to mobile phones' exterior, money, sound, and communication function while purchasing mobile phones, they are still concerned about the touch feel, week, and leisure. Besides, we do find that machine learning models can be used to predict the popularity and usefulness of online reviews accurately based on the number of stars, sentiment, and word counts. After combining the independent variables, the performance of the majority of machine learning models improves.

Some suggestions are made in this paper. Firstly, since the growing requirement of older adults for the multi-functions of mobile phones, manufacturers should make some adjustments to their innovation plans and spend some resources on exploring ways to improve the quality of leisure and work functions. Secondly, machine learning models have been successfully adopted to predict the popularity and usefulness of online reviews. Merchants can employ this method to locate useful comments in advance. If there are some problems mentioned in these useful reviews, merchants can respond in a timely manner.

References

Akca, M. Sayili, K.: Theory and practice: Esengun challenge of rural people to reduce digital divide in the globalized world. Gov. Inf. Q. **24**, 404–413 (2007)

Ball, C., Francis, J., Huang, K.T., Kadylak, T., Cotten, S.R., Rikard, R.V.: The physical-digital divide: exploring the social gap between digital natives and physical natives. J. Appl. Gerontol. **38**(8), 1167–1184 (2019)

Billon, M., Lopez, F.L., Marco, R.: Differences in digitalization levels: a multivariate analysis studying the global digital divide. Rev. World Econ. **146**, 39–73 (2020)

Cao, H.H.: Online review manipulation by asymmetrical firms: is a firm's manipulation of online reviews always detrimental to its competitor? Inf. Manag. **57**(6), 103244 (2020)

Chang, S.I., Yen, D.C., Chang, I.C.: Study of the digital divide evaluation model for government agencies - a Taiwanese local government's perspective. Inf. Syst. Front. **14**, 693–709 (2012)

Choi, J., Oh, S., Yoon, J., Lee, J.M., Coh, B.Y.: Identification of time-evolving product opportunities via social media mining. Technol. Forecast. Soc. Chang. **156**, 120045 (2020)

Cruz-Jesus, F., Oliveira, T., Bacao, F., Irani, Z.: Assessing the pattern between economic and digital development of countries. Inf. Syst. Front. **19**(4), 835–854 (2016). https://doi.org/10.1007/s10796-016-9634-1

Duplaga, M.: Digital divide among people with disabilities: analysis of data from a nationwide study for determinants of internet use and activities performed online. PLoS ONE **12**(6), e0179825–e0179825 (2017)

Estacio, E.V., Whittle, R., Protheroe, J.: The digital divide: examining socio-demographic factors associated with health literacy, access and use of internet to seek health information. J. Health Psychol. **24**(12), 1668–1675 (2019)

Fuchs, C.: The role of income inequality in a multivariate cross-national analysis of the digital divide. Soc. Sci. Comput. Rev. **27**(1), 41–58 (2009)

Ghasemaghaei, M., Hassanein, K., Benbasat, I.: Assessing the design choices for online recommendation agents for older adults: older does not always mean simpler information technology. MIS Q. **43**(1), 329–346 (2019)

Hall, A.K., Bernhardt, J.M., Dodd, V., Vollrath, M.W.: The digital health divide: evaluating online health information access and use among older adults. Health Educ. Behav. **42**(2), 202–209 (2015)

Huang, S.-C., Cox, J.L.: Establishing a social entrepreneurial system to bridge the digital divide for the poor: a case study for Taiwan. Univ. Access Inf. Soc. **15**(2), 219–236 (2014). https://doi.org/10.1007/s10209-014-0379-7

Jara, I., et al.: Understanding factors related to Chilean students' digital skills: a mixed methods analysis. Comput. Educ. **88**, 387–398 (2015)

Jiang, S., Hong, Y.A., Liu, P.L.: Trends of online patient-provider communication among cancer survivors from 2008 to 2017: a digital divide perspective. J. Cancer Surviv. **13**(2), 197–204 (2019). https://doi.org/10.1007/s11764-019-00742-4

Kiss, H., Fitzpatrick, K.M., Piko, B.F.: The digital divide: risk and protective factors and the differences in problematic use of digital devices among Hungarian youth. Child Youth Serv. Rev. **108**, 104612 (2020)

Li, M.X., Huang, P.: Assessing the product review helpfulness: affective-cognitive evaluation and the moderating effect of feedback mechanism. Inf. Manag. **57**, 103359 (2020)

Loo, B.P.Y., Ngan, Y.L.: Developing mobile telecommunications to narrow digital divide in developing countries? Some lessons from China. Telecommun Policy **36**, 888–900 (2012)

Loo, B.P.Y.: The e-society. Nova Science, New York (2012)

Manzuch, Z., Maceviciute, E.: Getting ready to reduce the digital divide: scenarios of Lithuanian public libraries. J. Am. Soc. Inf. Sci. **71**, 1205–1217 (2020)

Mudambi, S.M., Schuff, D.: What makes a helpful online review? a study of customer reviews on Amazon.com. MIS Q. **34**(1), 185 (2010)

Mumporeze, N., Prieler, M.: Gender digital divide in Rwanda: a qualitative analysis of socioeconomic factors. Telematics Inform. **34**, 1285–1293 (2017)

Neves, B.B., Waycott, J., Malta, S.: Old and afraid of new communication technologies? reconceptualising and contesting the 'age-based digital divide.' J. Sociol. **54**(2), 236–248 (2018)

Niehaves, B., Plattfaut, R.: Internet adoption by the elderly: employing IS technology acceptance theories for understanding the age-related digital divide. Euro. J. Inf. Syst. **23**, 708–726 (2014)

Olson, K.E., O'Brien, M.A., Rogers, W.A., Charness, N.: Diffusion of technology: frequency of use for younger and older adults. Ageing Int. **36**, 123–145 (2011)

Park, S.R., Choi, D.Y., Hong, P.: Club convergence and factors of digital divide across countries. Technol. Forecast. Soc. Chang. **96**, 92–100 (2015)

Petrovic, M., Bojkovic, N., Anic, I., Petrovic, D.: Benchmarking the digital divide using a multi-level outranking framework: evidence from EBRD countries of operation. Gov. Inf. Q. **29**, 597–607 (2012)

Potnis, D.: Inequalities creating economic barriers to owning mobile phones in India: Factors responsible for the gender digital divide. Inf. Dev. **32**(5), 1332–1342 (2016)

Reddick, C.G., Enriquez, R., Harris, R.J., Sharma, B.: Determinants of broadband access and affordability: an analysis of a community survey on the digital divide. Cities **106**, 102904 (2020)

Sachdeva, N., Tuikka, A.M., Kimppa, K.K., Suomi, R.: Digital disability divide in information society. J. Inf. Commun. Ethics Soc. **13**(3/4), 283–298 (2015)

Song, Z.Y., Wang, C., Bergmann, L.: China's prefectural digital divide: spatial analysis and multivariate determinants of ICT diffusion. Int. J. Inf. Manage. **52**, 102072 (2020)

Szeles, M.R.: New insights from a multilevel approach to the regional digital divide in the European Union. Telecommun. Policy **42**, 452–463 (2018)

Teso, E., Olmedilla, M., Martinez-Torres, M.R., Toral, S.L.: Application of text mining techniques to the analysis of discourse in eWOM communications from a gender perspective. Technol. Forecast. Soc. Chang. **129**, 131–142 (2018)

Tirunillai, S., Tellis, G.J.: Mining marketing meaning from online chatter: strategic brand analysis of big data using latent dirichlet allocation. J. Mark. Res. **51**(4), 463–479 (2014)

Townsend, L., Sathiaseelan, A., Gorry, F., Wallace, C.: Enhanced broadband access as a solution to the social and economic problems of the rural digital divide. Local Econ. **28**(6), 580–595 (2013)

Wang, J.N., Du, J.Z., Chiu, Y.L.: Can online user reviews be more helpful? evaluating and improving ranking approaches. Inf. Manag. **57**, 103281 (2020)