Journal of Korea Trade Vol. 24, No. 7, November 2020, 1-28
https://doi.org/10.35611/jkt.2020.24.7.1
ISSN 1229-828X

Global Big Data Analysis Exploring the Determinants of Application Ratings: Evidence from the Google Play Store

JKT 24(7)

Received 24 June 2020
Revised 20 August 2020
Accepted 7 October 2020

Min-Kyo Seo
Department of Trade and Management, Daegu University, South Korea

Oh-Suk Yang†
Division of Business Administration & Accounting, Kangwon National University, South Korea

Yoon-Ho Yang
School of Management, University College London, United Kingdom

Abstract

Purpose – This paper empirically investigates the predictors and main determinants of consumers' ratings of mobile applications in the Google Play Store. Using a linear and nonlinear model comparison to identify the function of users' review, in determining application rating across countries, this study estimates the direct effects of users' reviews on the application rating. In addition, extending our modelling into a sentimental analysis, this paper also aims to explore the effects of review polarity and subjectivity on the application rating, followed by an examination of the moderating effect of user reviews on the polarity-rating and subjectivity-rating relationships.

Design/methodology – Our empirical model considers nonlinear association as well as linear causality between features and targets. This study employs competing theoretical frameworks – multiple regression, decision-tree and neural network models – to identify the predictors and main determinants of app ratings, using data from the Google Play Store. Using a cross-validation method, our analysis investigates the direct and moderating effects of predictors and main determinants of application ratings in a global app market.

Findings – The main findings of this study can be summarized as follows: the number of user's review is positively associated with the ratings of a given app and it positively moderates the polarity-rating relationship. Applying the review polarity measured by a sentimental analysis to the modelling, it was found that the polarity is not significantly associated with the rating. This result best applies to the function of both positive and negative reviews in playing a word-of-mouth role, as well as serving as a channel for communication, leading to product innovation.

Originality/value – Applying a proxy measured by binomial figures, previous studies have predominantly focused on positive and negative sentiment in examining the determinants of app ratings, assuming that they are significantly associated. Given the constraints to measurement of sentiment in current research, this paper employs sentimental analysis to measure the real integer for users' polarity and subjectivity. This paper also seeks to compare the suitability of three distinct models – linear regression, decision-tree and neural network models. Although a comparison between methodologies has long been considered important to the empirical approach, it has hitherto been underexplored in studies on the app market.

Keywords: Application, App Store Market, Big Data, Cross-Validation, Decision-Tree, Neural Network Model, Sensitivity Analysis

JEL Classifications: C45, D12, M16

† **Corresponding author:** osyang30@kangwon.ac.kr

1. Introduction

We are in the age of the "customer experience", when not only quality but also the customer's experience in purchasing and consuming a product are increasingly important. "Customer delight" has become a key predictor of firm performance, going beyond the scope of the traditional "customer satisfaction". Various marketing strategies that look to maximize "customer delight"—such as relationship marketing, emotional marketing and experience marketing—have been discussed in the literature or implemented in practice by firms. These days, managing the customer's experience is necessary to gain a distinctive competency and create long-term value.

Marketers are increasingly paying attention to the power of word-of-mouth (WOM) in promoting customer awareness and building brand identity (Anderson and Sullivan, 1993; Brown and Reingen, 1987; Herr, Kardes and Kim, 1991; Leonard-Barton, 1985; Liu, 2006; Oliver and DeSarbo, 1988; Schmitt, Skiera and den Bulte, 2011). In the online market, customers' online reviews play this same role (Baker and Algorta, 2016; Berger and Schwartz, 2011; Chevalier and Mayzlin, 2006; Lee Eun-Ju and Shin Soo-Yun, 2014; Lovett, Peres and Shachar, 2013). Online reviews comprise two elements – the objective experience of the customer and their consumption emotions (Phillips, 1999). Such emotions can span the spectrum but for analytical purposes they are typically categorized as either positive or negative, which may be the aggregate of different emotions (Oliver, 1997; Phillips, 1999). They determine the level of satisfaction or dissatisfaction (Oliver, 1997) and prompt subsequent consumer behavior such as WOM or complaints, affecting the firms' profits (Oliver, 1997).

The extant research into consumption emotions is broad and varied. Some researchers have investigated the credibility, persuasiveness and receptivity of reviews (Lee Kook-Yong, 2017), while others have examined their efficacy in affecting other consumers' purchases, proposing that this relationship is mediated by the customer's attitude and flow (Baier and Stüber, 2010; Gupta and Harris, 2010; Hausman and Siekpe, 2009). Other researchers have considered the relationship between the credibility and quality of reviews and the consumer decision-making process (Filieri, 2016; Pavlou and Gefen, 2004; Ponte, Carvajal-Trujillo and Escobar-Rodriguez, 2015). Studies on the relationship between the credibility and expertise of the review source and the customers' acceptance of information represents another strand of work (Cheung and Lee, 2008; Filieri, 2016; Zhang and Watts, 2008). Other studies have compared the influence on prospective customers of reviews by average consumers and experts (Lee Dong-Il and Choi Seung-Hoon, 2012), while others have looked at the impact of the number and polarity of reviews on firm performance (Berger, Sorensen and Rasmussen, 2010; Lee Dong-Il and Choi Seung-Hoon, 2012; Zhu and Zhang, 2010).

Such studies reveal two disjunctions. Firstly, most previous studies (Berger, Sorensen and Rasmussen, 2010; Lee Dong-Il and Choi Seung-Hoon, 2012; Zhu and Zhang, 2010) have regarded reviews simply as an online form of WOM, focusing on the linear relationship between with firm's financial performance or ratings, i.e. firm's market performance. In doing so, they have not exploited the methodological advantages online reviews offer, such as their significant out-of-sample prediction power, which has been provided by other supervised learning tools, including decision-tree and neural network models based on cross-validation (Franses and Van Griensven, 1998; Kuan and Liu, 1995; Swanson and White, 1995; Lee Sang-Jae and Choeh Joon-Yeon, 2014; Phillips et al., 2015). The second disjunction is that the majority of studies have focused only on the size, polarity and quality of reviews due to a lock-in to the motivation-emotion-outcome model. This approach has caused researchers to disregard leading or trailing relationships between the size and polarity of reviews on the one hand and ratings on the other. Moreover, the moderating function of reviews for the emotion-outcome relationship, too, has been overlooked.

This study aims to open the black box in which the latent relationship between the number and polarity of reviews operates. This study also seeks to verify how this relationship affects ratings and to deliver an advanced insight regarding the emotion-behavior-outcome association. In this paper, using sentimental analysis, we identify the determinants of app market performance, as measured by its rating, in order to improve app quality, and thereby firm performance. Although this paper considers only a single case, the role of the reviews has been rigorously verified through the use and comparison of various models, each with different predictive powers, such as regression, decision-tree and neural network models.

Our results show that, in contrast to the observation that apps with a greater number of reviews tend to have higher ratings, whether the reviews carry positive or negative emotion is irrelevant to the overall rating. This implies that it is not the emotional propensity that determines an app's rating, but rather the behavioral intention such as reviews moderates the rating. Therefore, in the users' rating function, behavioral factors are more important than perceptual or attitudinal factors.

This article is organized as follows: the introduction sets out the research question, and our hypothesis is put forward in Chapter 2. Chapter 3 explains our methodology, including our research model, data sampling methods and descriptive statistics. This is followed in Chapter 4 by an analysis of both the results and their validity. In Chapter 5, we explore the implications of these results, in terms of theory and practical business management, before considering the limitations of this study and further research ideas in Chapter 6.

2. Theoretical Background and Hypotheses

Traditionally, in Marketing, the customers' purchasing motivation is understood as affecting the purchasing emotion, which in turn leads to behavioral outcomes such as purchasing decisions, customer loyalty, purchasing preferences, repeat purchases and word-of-mouth. This is based on the Motive-Emotion-Outcome (MEO) model of Dawson, Bloch and Ridgway (1990). This model has been used in many studies to elucidate the relationship between consumers' motive-emotion-outcome, and in doing so has been modified and developed by several scholars (Babin, Darden and Griffin, 1994; Childer et al., 2001; Hammond, McWilliam and Diaz, 1998).

This paper seeks to revise and supplement the MEO model to overcome its inherent limitations. Firstly, this model. The first methodological divergence is that this model divides the consumer's emotions into seven: relaxed, content, satisfied, happy, surprised, excited and rewarded (Dawson, Bloch and Ridgway, 1990, 417). This classification of emotions has been further subdivided by a number of scholars (Oliver, 1993; Swinyard, 1993; Yoo Chang-Jo, Park Jong-Hee and Maclnnis, 1998). For instance, Hoffman and Novak (1996) coined "flow", defined as a positive empirical process achieved by users interacting with a computer-mediated environment, while maintaining a balance between skill and challenge. In recent years, app users interacting with a computer-mediated environment and seeking information-oriented aspects, enjoy the search for information itself, and interact with each other through reviews. This study aims to measure the emotions of modern-day consumers at a multi-category level, using a deep learning-based sentiment analysis tool which divides all emotions into simplified positives and negatives.

The second methodological divergence is that the original MEO model treats the interaction between Emotion and Outcome as a black box. Dealing with this constraint, this study attempts to interrogate this relationship by examining consumer reviews as a link between the two and considering their mediating and moderating functions. In particular, it is expected that the significance of this effect might differ between retailers in markets with products of homogeneous quality, such as gaming apps, and those of heterogeneous quality, such as agricultural products.

There are many different metrics used to predict the success or failure of an app. These include profitability, the number of downloads, growth rate and the app store rating. For instance, the Google Play app store ranks apps by the number of downloads, as well as the amount of "max profit" or "sudden increase" in order to display the popularity of its apps. One common feature of highly ranked apps is that they tend to have high average user rating. Therefore, it is expected that the popularity of an app would correlate with its rating, which has been found in turn to be proportional to the number of downloads.

Any user who decides to download an app faces potential risks, along with the reliability of the market (Gefen, Rao and Tractinsky, 2003; Kim Gi-Mun and Koo Hoon-Young, 2016). These aspects of the app market affect the decisions of users in different ways. According to Kim Gi-Mun and Koo Hoon-Young (2016), potential risks in the market do not affect consumers' intentions with respect to downloading apps, while the reliability of the market does. Reliability is considered to be an essential element in the majority of online transactions (Kim, Ferrin and Rao, 2008; Pavlou, 2003). Therefore, the reliability of the app market is expected to be an important factor in determining whether consumers decide to download a specific app or not.

The reliability of a particular app is not easily built immediately after a customer downloads an app. Customers tend to more readily trust apps that have positive reviews and/or a large number of users. The app market ranks apps by the number of downloads. A large number of downloads may imply that many users have confirmed the quality of the app. Additionally, users tend to believe that reviews of previous users are thorough and reliable, which makes new users trust more popular apps (Gu et al., 2017). Apps with high download counts are therefore generally considered to be more reliable by first-time users because they believe the high number of downloads is a form of endorsement. This was proven as the majority of users download the most popular apps (Zhong and Michahelles, 2013).

Apps with high download counts are also likely to have a large numbers of users, allowing them to create vast community platforms for the sharing of information about app. Large-scale app community platforms are more effective for communication between the user and the developer than smaller ones. Effective communication between the developer and the users of the app also allows the developer to produce regular updates to improve the quality of the app, which ultimately leads to a high level of customer satisfaction (Frie et al., 2017). These factors are expected in turn to increase the rating of the app.

In conclusion, a high download count, which corresponds to app popularity, signifies high reliability and motivates downloads. Such apps also enjoy large-scale online communities, providing a channel for the delivery of feedback from users to the developer, which produces a virtuous circle, with a positive impact on the app rating. This idea is encapsulated in the hypothesis below.

H1: The more times an app has been downloaded, the higher the rating of the app.

A high rating for a specific app should not be considered merely a quantitative measure of popularity, but rather a combination of the process and factors. In other words, there are many aspects that determine the rating of an app, with results following black-box dynamics. There are a variety of factors that affect a consumer's evaluation of a given app. Previous studies have shown that system requirements, app size and the accessibility of the user interface are important factors for users' assessments of app quality (Frie et al., 2017; Yuan, 2015). Such aspects of apps are frequently referenced in reviews. In this way, written reviews reveal the factors that are important to users in determining app quality.

App reviews provide those considering downloading the app with further information beyond the app description written by the developer (Bardus and van Beurden, 2016). Existing reviews may impact subsequent ratings, as users' opinions may be influenced by the reviews they read. A previous study has shown that apps with a large number of useful reviews tend to experience an increase in rating during distribution (Palomba and Linares-Vasquez, 2015). This finding was derived from an analysis of the number of reviews having already shown the positive correlation between the number of reviews and the rating of an app.

App reviews provide a channel through feedback can be delivered to developers (Iacob and Harrison, 2016). Having many reviews may indicate that there is a frequent exchange between users and developers, and that issues highlighted receive a response, perhaps by changes to the app in updates. It has been found that the more regularly an app is updated, the higher the app's ranking in the market (Cho Hyu-Kjun, Kang Ju-Young and Jeong Dae-Yong, 2016).

In summary, the greater the number of reviews an app has, the greater the chance that customer feedback will be reflected in improving the quality of the product. Additionally, reviews may include suggestions and which can be utilized by developers to improve the app to satisfy the needs of the customers (Frie et al., 2017), further increasing the chance of higher ratings. The following hypothesis is put forward.

H2: The greater the number of written reviews an app has, the higher the overall rating.

There are several motivators and indicators of consumerism in the app market. Users generally consult the reviews, ratings, price and download count of the application before making a decision on whether or not they should use the app. Even at this browsing stage, before even using the app, consumers establish an impression of the app. The question that arises is if there are implications for rating in a review written by a user who had already established a premature impression on the app based on the reviews. In other words, are a user's positive or negative impressions of an app a function of the existing average rating of the app?

A previous study into consumer behavior found that a more positive impression of an app by other users will lead to the giving of a more favorable rating (Martens and Johann, 2017). However, there were many outliers in the study which weakened the correlation observed. Hence, the question in which whether the extent of the impressions of the customer affected their level of satisfaction remain unanswered. Nonetheless, another study observed that the tone of written reviews correlated significantly with the average rating of the app (Hoon, Vasa and Schneider, 2013). The same study noted that the correlation between the positivity of a review and a higher average rating was particularly strong, reaching a probability of 95% (Hoon, Vasa and Schneider, 2013).

By contrast, this paper argues that as perceptional and attitudinal variables, the opinions of users – whether they be positive or negative – do not significantly affect the overall rating of an app, which is a resultant variable. Unlike those factors, as referred to in Hypothesis 2, behavioral factors such as writing reviews yield significant results on the average rating of an app. The act of writing and submitting a review of an app is a "behavior", while the opinions and emotions contained therein are "perceptions" and "attitudes". Written reviews can be described as tangible, while opinions or impressions are not. Applying this concept to the app market, intangible cues, such as the opinions of users, are not directly associated with the rating, while tangible cues, such as user reviews, are more likely to have a direct effect on brand image or the app rating.

In conclusion, the assessment of an app is not determined by a user's positive or negative response to it, but rather by the existing reviews published in the market. This process allows customers to relay their positive or negative opinions about the app through the moderating of reviews, which guides the developers in making improvements. Through this user-developer interaction, it can be concluded that the relationship between user opinions and the average app rating is statistically significant. The following hypotheses are proposed.

H3: There will be no significant relationship between the positive and negative opinions of users towards the app and the app rating.

H4: Reviews will have a positive impact on the relationship between positive and negative opinions of users and the app rating.

Consumers have the basic purpose of satisfying their needs in product consumption. Consumer desires, which act as purchase motivations, consist of functional, symbolic, and experiential consumption values or benefits (Aaker, 1995; Keller, 1993; Park, Jaworski and Maclnnis, 1986). Firstly, functional benefits related to the intrinsic and specific properties of a product or service (Keller, 1993) include functional motives based on product information (Westbrook and Black, 1985). Meanwhile, experiential benefits (Keller, 1993) – i.e. the emotional and cognitive stimulation felt by the consumers while using products or services – encompass the pursuit of pleasant experiences or experiential motivations based on previous experiences with the same product (Westbrook and Black, 1985). Finally, symbolic benefits related to the role and status of consumers, a sense of belonging to a certain group and self-identification (Keller, 1993) can be divided into social motivations such as communication with fellow consumers, the pursuit of status and power, and incentives of reference groups (Tauber, 1972).

The price of an app, perceived as a prerequisite for app ratings competing with emotion, is linked to motivation. Price is one of the functional benefits of consumers and can be specifically classified as an economic benefit. When a user considers downloading a particular app, will economic benefits such as price or motivation act as a new added feature? There remains no consensus among researchers on the answer to this question. For instance, Harman, Jia and Zhang (2012) found that price and downloads were not related, and Pagano and Maalej (2013) observed that the volume of feedback increased regardless of the price of the application. According to these previous studies, consumers' behavior such as giving feedback is based on intrinsic motivation. In the same vein, it is expected that price will not be a factor in determining app ratings.

H5: There will be no significant relationship between price and the app rating.

Given the fact that human emotions are subjective, it is necessary to consider subjectivity as another important aspect of reviews. Subjectivity reflects the beliefs, attitudes and values each individual holds about a specific object or issue (Brown, 1980). In short, subjectivity is a deviation that shows to what extent someone's evaluation of an application differs to those of others (Liu, 2010; Nayebi, Cho and Ruhe, 2018). The content of reviews on specific app experiences may reveal either very strong or insignificant positive and negative attributes depending on the strength of the individual consumer's subjectivity. The problem is that consumers tend to post reviews with extreme subjectivity rather than moderate. In addition, consumers who have undergone extremely positive or negative experiences will be more inclined to share this (Dellarocas, Awad and Zhang, 2005).

It is, therefore, noteworthy that reviews with high subjectivity tend to reinforce subjective information. These reviews lose their utility and validity, while at the same time reducing consumer trust and hindering the original function of reviews in facilitating communication with developers. This means that subjectivity modulates the relationship between reviews and ratings – in effect, the relationship between reviews and ratings can either be negatively adjusted or be decoupled. In this article, based on the latter perspective, which assumes the maximization of the influence of subjectivity, the hypothesis is made that the interaction term between review and subjectivity has no statistically significant relationship to the rating.

H6a: There will be no significant relationship between subjectivity and the app rating.

H6b: Subjectivity will have no significant impact on the relationship between reviews and the app rating.

3. Empirical Method and Data

3.1. Research Model

Prior to analysis of the app data, a word cloud was generated using text extracted from Google Play reviews (n=6,560). A word cloud is a graphic representation of text data that gives greater prominence – in terms of font size and placement – to words that appear more frequently in the sample. It was anticipated that this would give an overview of the current market situation, indicate trends and suggest a potential direction for this investigation. “Game” was the most common word, suggesting game apps attract the greatest proportion of reviews. This was followed by positive emotive terms such as “love”, “like”, “good” and “great”. This could be interpreted as suggesting that a strong emotional response to an app may motivate users to submit a review. This paper will investigate this relationship using the following regression model:

$$Y=ß_0+ß_1X_1+ß_2X_2+ß_3X_3+ß_4X_4+ß_5X_5+ß_6X_6...ß_9X_9+ß_7X_1X_3+e \quad (1)$$

where *Y (Rating),* X_1 *(Review),* X_2 *(Install),* X_3 *(Polarity),* X_4 *(Subjectivity),* X_5 *(Price),* $X_6 \sim X_9$ *(Dummy Variables: Category, Size, Type and Content.Rating),* X_1X_3 *(Interaction term: Review*Polarity) and* ε *(error term).*

3.2. Measure

Collecting the basic data consisting of 10 times, we calculated *Polarity (Sentiment)* and *Subjectivity* as additional variables. Employing sentimental analysis, whether the emotional content of the review was either positive or negative in tone was determined and assigned a *polarity* value using the SentiWordNet dictionary. This had a range of 0 as a minimum and 0.93 as a maximum. *Subjectivity* indicates the deviation of an individual user’s positive or negative assessment from those of other reviewers, with a range of 0.08 to 0.91.

The dependent variable, *rating,* measured the level of users’ cognitive assessment over the last three years using a five-point Likert scale. Firm performance has typically been measured by three metrics – financial performance, market performance and innovative performance; this study took the ratings of a firm’s apps as the leading indicator of financial performance.

Fig. 1. Word Cloud of Key Words

As a predictor and a moderator, *review* was the number of reviews for the given app, showing a broad scalar. As another predictor, *install* was the number of unique installations of the app, also showing a broad scalar. *Price* was a predictor, ranging from free to a maximum of 79 USD.

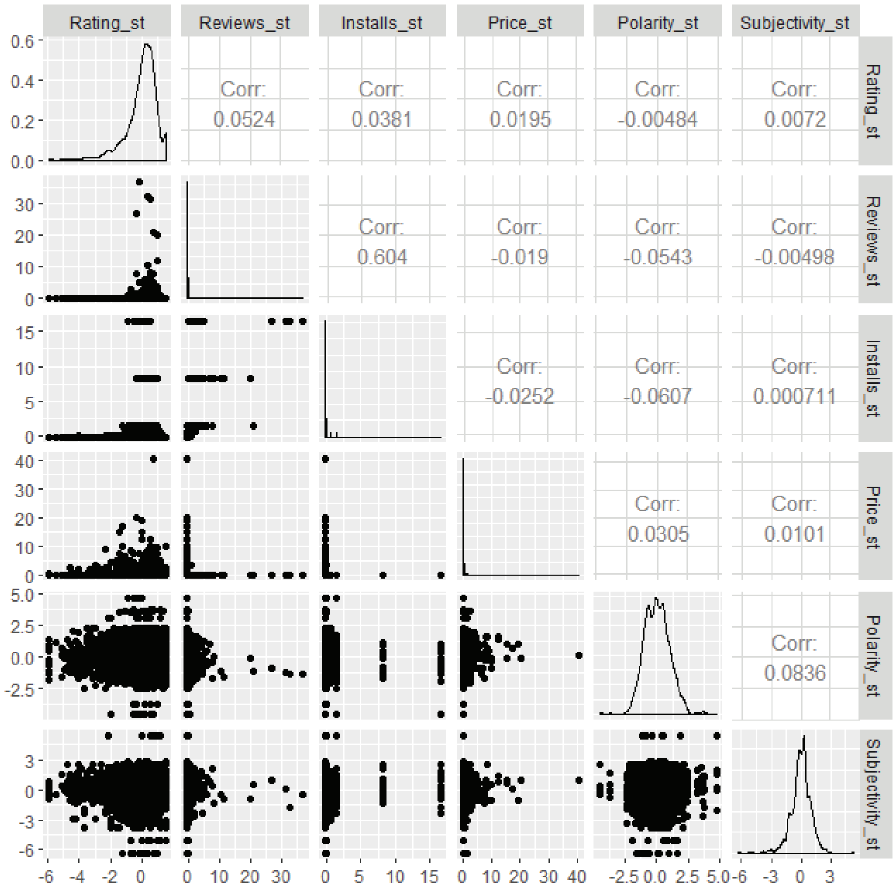
As control variables, *category, size, type and content.rating* were dummy variables. Apps within the data set spanned a total of 32 categories, including Family (1284 observations), Game (730), Tools (575), Productivity (241), Personalization (239), Finance (237) and Other (3254). In terms of *Size*, apps were considered to be either Big (2764) or Small (2863). *Type* divided apps into either Free (6103) or Paid (457), while apps had one of six values for *content.rating*, i.e. Adults 18+ (3), Everyone (5294.), Everyone 10+ (257), Mature 17+ (276), Teen (729) and Unrated (1). Finally, to confirm the suspected interaction between *review* and *polarity*, an interaction term, *review*polarity*, formed part of the study.

3.3. Data Collection and Description

Google launched its official app store, Google Play, in 2008. It offers a vast range of applications – now numbering over 2.6 million – as well as digital media, including music, magazines, films and TV shows. With a gross revenue of $24.8 billion, it is the largest and most popular app store for Android devices. This study collected data on application rating and review from online channels by crawling and used a total set of 6,560 reviews submitted to the Play store from 2012 to 2018. 10 pieces of data were collected for each app: *rating, app, category, review, size, install, type, price, content.rating* and *lastupdated.*

Applying the cross-validation method, we calculated the mean of the Sentiment Polarity and Subjectivity columns and then merged these processed columns with data from the app store. NA (not available) values were generated afterwards and we input them, making use of a Random Forest-like prediction model. Similar methods were used for the test data set.

Fig. 2. General Information of Variables



The data set comprised 6,560 observations and 12 selected variables (Rating, App, Category, Review, Size, Install, Price, ContentRating, LastUpdated, Polarity, Subjectivity). The general plot of the data set is shown in Fig. 2. This graph shows the relationship between each pair (n=6), and the general relationship and dispersion of the observations. From this graph, it was possible to confirm a correlation between the variables and the respective normal distributions. For example, rating was skewed to the right, implying the existence of outliers, which are investigated later in this paper. It was confirmed that there are outliers within the data for some variables, but we decided against controlling for them since it was considered that these outliers might represent meaningful observations.

Aiming at realizing fundamentally the normality of the data set, this paper standardized each variable such that it had a mean of 0 and a standard deviation of 1 (see Table 1). However, in the case of *rating* the skew of all variables came out as negative, confirming that the distribution was biased to the right compared to a normal distribution. The distribution quartile of the variables was also analyzed. Significant outliers were considered to be those exceeding three times the interquartile range (IQR) (Outlier > UpperQ+IQR*1.5 or <

LowerQ-IQR*1.5): out of 6,560 observations there were six significant outliers found below the bottom quartile for *review*, two for *install* and one for *price*. Other variables had no significant outliers. These were considered negligible and so it was decided not to control for them.

Fig. 3. Outlier by Inter-Quartile Range

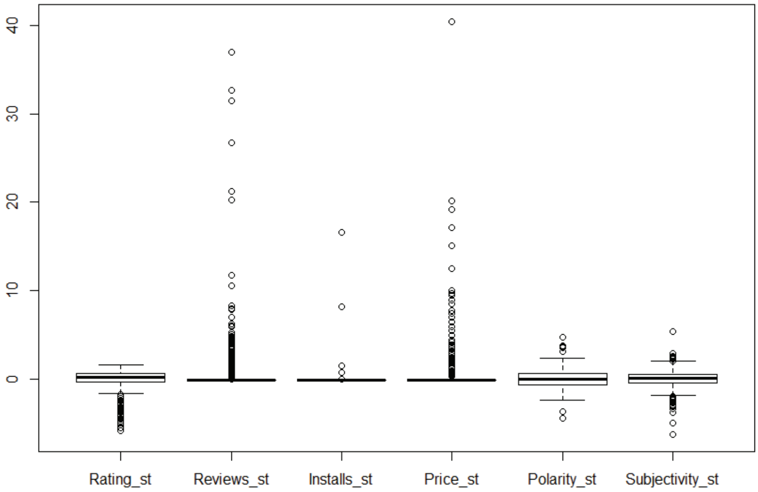


Table 1. Descriptive Statistics (standardized)

Variable	Mean (standardized)	Standard Deviation (standardized)	Max	Min	Outlier
Review	262060(9.000)	2105226(1)	37.001	-0.124	6
Install	0.001(0.000)	5980929(1)	16.562	-0.157	2
Price	0.317(0.000)	1.968(1)	40.483	-0.161	1
Polarity	0.198(0.000)	0.155(1)	4.723	-4.486	5
Subjectivity	0.496(0.000)	0.078(1)	5.346	-6.312	3
Rating	4.173(0.000)	0.537(1)	1.537	-5.901	0

Note: The figures in brackets are standardized values.

With regards to descriptive statistics, the correlations, means, standard deviations and VIF (variance inflation factor) results are presented in Table 1. There were no VIF values higher than 0.7, the conventional threshold for determining the presence of collinearity between variables (Anderson, Sweeney and Williams, 1996; Griffiths, Hill and Judge, 1993). The second check employed the VIF function on the basis that the coefficient of VIF increases when there is a strong correlation between supposedly independent variables. Again, by convention, collinearity is indicated by a VIF value of 10 or above (Chatterjee and Price, 1991; Hair et al., 1998; Kennedy, 1992), while at times a stricter standard of 3.3 is applied (Diamantopoulos and Siguaw, 2006; Petter, Straub and Rai, 2007). All values in this experiment were much smaller than either of these threshold figures, implying again that there were no issues with multicollinearity.

Table 2. Collinearity and VIF

Variable	(1)	(2)	(3)	(4)	(5)	VIF
Install	1					1.573
Price	0.603	1				1.575
Polarity	-0.019	-0.025*	1			1.001
Subjectivity	-0.054*	-0.060*	0.031	1		1.012
Rating	-0.004	0.001	0.010	0.083*	1	1.007

Notes: 1. The figures in brackets are standardized values.
2. *p<0.01.

3.4. T-test

It is not necessary to take a unit root test as the data in this research is not that of a time series. Instead, since the hypothesis concerns a comparison between positive and negative groups, the main analysis was preceded by a T-test to check for a difference in their variances. The normality test was conducted to confirm the normal distribution of the data set and then a T-test was performed through a standard F-test and Bartlett test. This was followed by a nonparametric Ansari-Bradley test – this verifies the suitability of a standard independent samples T-test. As a result of the F-test and Bartlett test, it was confirmed that variances of the positive and negative groups were different and that there was no overlapping confidence interval. That is, the two groups do not have the same mean and therefore we can conclude that the positive and negative groups evaluated apps sufficiently differently.

4. Empirical Results

4.1. Method: Model Comparison

This study adopts a cross-validation approach to examine the efficiency of three predictive models—a regression model, a decision tree model and a neural network model—in analyzing the factors that influence app ratings.

In contrast to the basic regression model, the decision tree model shows multi-scale, complicated conditions and presents solutions in a tree form. The largest condition is represented by the root of the tree, from which branches give more detailed conditions. The decision tree model, which is based on a chart representation of the decision rule model that categorizes and predicts outcomes, can itself be used to make predictions but is also a form of non-linear regression or non-linear discriminant analysis and is used for classification. Additionally, the decision tree model can be used to select the factors affecting a target variable, especially in the presence of multiple variables. Therefore, it is often used as a means of exploring a research question before other regression techniques are pursued (Choi Jong-Hoo, 2000). The decision tree model produces visual results and has the advantage of being simpler to understand over the regression or neural network models, making it the best suited for explaining an analytical process, rather than necessarily producing the most accurate results (Choi Jong-Hoo, 2000).

The neural network model was first developed in the field of computer science in the early 1980s for solving problems in the natural sciences. However, throughout the 1990s, statistical analysis became an active area of research, with the development of non-linear regression, discriminant analysis and projection pursuit regression (Kim Sang-Hwan, 2000; Ripley, 1993;

Weiss and Kulikowski, 1991), and the neural network model found use there, too. It is generally assumed that the neural network model approximates non-linear regression functions more accurately than the random-walk model (Kim Sang-Hwan, 2000). The model is especially useful for "multilayer perceptrons", in which multiple hidden layers exist between the input and output layers and where there is no information that links the input and output variables (Barron, 1993). Most papers on this model agree that it is superior thanks to its ability to predict outcomes with a high degree of accuracy (Franses and Van Griensven, 1998; Kuan and Liu, 1995; Swanson and White, 1995). However, its disadvantages include having multiple local minima and its overfitting of noise in data, making it sensitive to such noise when making predictions (Kim Sang-Hwan, 2000). Hence, Kim Sang-Hwan (2000) concludes that, despite the advantages of the neural network model, proper model selection is essential for improving the prediction accuracy of the model.

In summary, each statistical model has distinct advantages and disadvantages, such that each is best suited to different types of data. This paper compares the predictive and generalization powers of these three models to find which is most appropriate for explaining the factors that determine the rating that users assign to apps. Models should be capable of processing out-of-sample predictions, as well as explaining training data –– the cross-validation method is normally used to assess this. The approach involves a training dataset of function *f* that shows the relationship between the explanatory variable and the dependent variable, and a process that checks the selected model to assess the predictive power by using a divided test data set. In this study, adopting SEMMA (Sampling, Explore, Modify, Model, Assess) approach, the training set and the test set were divided by a ratio of 5:5, that is, each divided into 3,280 to confirm the training error rate and test error rate. Then we implemented the selection of input variables needed to identify the model with the smallest error rate. For this purpose, the method of Helsel and Hirsch (1992/2002), which evaluates the combination of independent variables for models consisting of all possible combinations of independent variables, was adopted, and the result was the same as the regression equation (1) presented above.

The cross-validation method is superior to other methods that use a large data set, such as Akaike (1973)'s final prediction error or Moody (1994)'s generalized prediction error criterion, in that the cross-validation method searches for the optimum model regardless of the size of the data set (Kim Sang-Hwan, 2000). Concerning model accuracy statistics, under regression setting, RMSE and MSE are the most-frequently-used measurement. The three models will be assessed in terms of their predictive power of out-of-sample data using the mean square error and root mean squared error, which place greater weight on larger errors (Frechtling, 2001; Witt and Witt, 1995) (see Table 5). Lift Charts and Receiver Operating Characteristic charts will also be used (Berry and Linoff, 1997).

4.2. Result of Multiple Linear Regression Model

4.2.1. Multiple Linear Regression

The results of all models showed that "Review" was proportional to "Rating". This was consistent with the outcome of the prediction from the regression tree. Therefore, the sales performance of an app is influenced by the number of user reviews. In contrast, the relationship between "Rating" and each of "Install", "Price" and "Content Rating" was not significant in any of the three models used. With regards to factor variables, "type paid" (Model 4, 6, 8) and size (varies with device) (Model 3, 6, 8) were proportional to rating, too. Notably, "Size (small)" was negatively associated with rating, and only in model 3. Meanwhile, the statistical significance of "Category" was partially and clearly distinct in all models,

Table 3. Result for Multiple Linear Regression

Hypotheses	Variables		Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
H2	Predictor	Review	0.046***	0.041***	0.044***	0.047***	0.045***	0.039***	0.137***	0.124***	0.042***
H1	Indicators	Install	0.011	0.014	0.001	0.012	0.009	0.005	-0.001	-0.005	0.005
H5		Price	0.021	0.020	0.021	-0.017	0.021	-0.012	0.022	-0.012	-0.012
H3		Polarity	-0.003	-0.003	-0.002	-0.003	-0.003	-0.003	0.001	0.001	-0.003
H6a		Subjectivity	0.007	0.006	0.008	0.008	0.008	0.007	0.007	0.007	0.007
	Control Variables	Category		Partial				Partial		Partial	Partial
		Size (small)			-0.066*			-0.021		-0.018	-0.021
		Size (varies)			0.111**			0.146***		0.138***	0.146***
		Type (Paid)				0.252***		0.218***		0.225***	0.218***
		Content Rating					No (All)	No (All)		No (All)	No (All)
H4	Moderator	Polarity*Review							0.089***	0.084***	
H6b	Moderator	Review*Subjectivity									0.001
		Intercept	0.000	0.253	0.013	-0.018	0.244	0.539	0.005	0.550	0.539
		Adjusted R-Squared	0.003	0.026	0.006	0.005	0.004	0.029	0.005	0.032	0.029
		F-statistic	4.340†	5.681†	6.294†	6.581†	3.497†	5.468†	6.788†	5.723†	5.353†

Notes: 1. Dependent Variable: Market Capability.
2. "Partial" indicates some categories are significantly associated with dependent variable; "No (All)" means all contents do not have any significant association with dependent variable.
3. *p<0.05, **p<0.01, ***p<0.001, †p<0.01.

suggesting it may influence rating through various means. "Polarity" and "Subjectivity" showed a statistical significance in all models. Most interestingly, "Polarity" in itself did not show statistical significance, but became statistically significant through the interaction term between it and "Review".

4.2.2. Robustness Test

This paper used four different types of robustness test to obtain consistent results. First, we input control variables in models in order instead of inputting them. The results showed that the coefficients were slightly different, while the significance and the direction of the relationship were unchanged. Second, the Wooldridge test was applied to verify the existence of first-order autocorrelation for an error term, which panel data may involve (Wooldridge, 2002). For all models, the value of Prob > F did not lead to the rejection of the null hypothesis – "there is no first-order autocorrelation" – proving the absence of first-order autocorrelation. Third, we examined the Granger causality between "Review" and "Rating" by using a VAR model. Due to the non-significance of the p values across all models, the null hypothesis – "Review does not Granger-cause Install, Price, Polarity, Subjectivity, Rating and Interaction term" – was accepted. Finally, to confirm the existence of other moderators, this paper applied interaction terms in each model. It was found that no other interaction terms had moderating effects. The mediating effect of "Review" on the relationship between positive or negative feelings and rating was also confirmed, and found to be not significant.

Table 4. Robustness Test

Variables		**Model 10**	**Model11**	**Model12**	**Model13**	**Model14**
Predictor	Review *H2*	0.039***	0.039***	0.039***	0.042***	0.040***
Indicators	Install *H1*	0.005	0.005	0.005	0.011	0.005
	Price *H5*	-0.012	-0.001	-0.013	-0.012	-0.012
	Polarity *H3*	-0.003	-0.003	-0.002	-0.003	-0.003
	Subjectivity *H6a*	0.007	0.007	0.007	0.007	0.007
Control Variables	Category	Partial	Partial	Partial	Partial	Partial
	Size (small)	-0.021	-0.021	-0.021	-0.021	-0.021
	Size (varies)	0.145***	0.148***	0.146***	0.145***	0.146***
	Type (Paid)	0.218***	0.207***	0.219***	0.218***	0.218***
	Content Rating	No (All)	No (All)	No (All)	No (All)	No (All)
Moderator	Polarity*Subjectivity	0.006				
	Polarity*Price		-0.035*			
	Subjectivity*Price			-0.035*		
	Polarity*Install				0.012	
	Subjectivity*Install					0.001
Intercept		0.541	0.538	0.539	0.541	0.539
Adjusted R-Squared		0.029	0.030	0.029	0.029	0.029
F-statistic		5.355†	5.442†	5.348†	5.377†	5.348†

Notes: 1. Dependent Variable: Market Capability.
2. "Partial" indicates some categories are significantly associated with dependent variable; "No (All)" means all contents do not have any significant association with dependent variable.
3. *p<0.05, **p<0.01, ***p<0.001, †p<0.01.

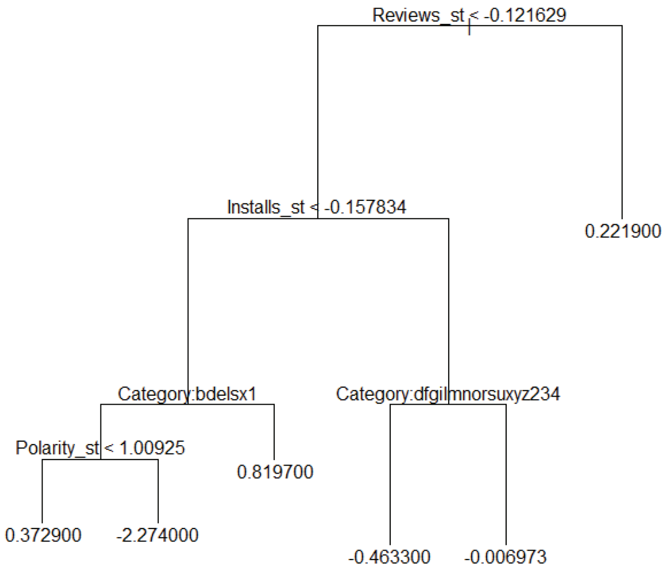
4.3. Result of Tree Model

4.3.1. Regression Tree

Tree-based models are simple and effective to interpret. Our final selection for the model of regression tree included the interaction term, *Polarity x Review*. Using a tree plot, it was determined that “Reviews” and “Installs” have a significant relationship with the dependent variable (“Rating”). The regression tree in Fig. 4 has four internal nodes (Review, Install, Category and Polarity), and a split that produces two main branches. The left-hand branch corresponds to Reviews < -0.12, while the right-hand branch relates to Review >= -0.12. The number in each leaf (or terminal node) is the mean of the response for the observations that fall into that subgroup (James et al., 2013, 304).

“Review” is the most important determinant of “Rating”, such that apps with fewer reviews achieve a lower rating than those with a greater number. Given that an app is more reviewed, the number of times the app has been installed seems to have little impact on in his or her rating. However, among apps with more reviews, the number of times an app has been installed does affect the rating – apps that have been installed a greater number of times tend to have higher rating. Finally, using cross-validation with a scaled test dataset, we obtained an RMSE of 1.0060.

Fig. 4. Regression Tree



4.3.2. Robustness Test

The Random Forest Tree (RFT) model was also conducted as a robustness test. This model complements the regression tree and bagged tree models by selecting a predictor subset size and pursuing “m=√p” (where p = the number of predictor variables) (James et al., 2013, 319-321). Although the RFT and bagged tree models have the same method, a smaller boosting factor is appropriate for the RFT. Furthermore, it produces trees with higher accuracy, so it provides more functions than the bagging tree model (James et al., 2013, 319-321). When applying the RFT model, we included all the given variables and the interaction term (*Polarity x Review*) and used a cross-validation procedure on a scaled test dataset. We obtained an RMSE of 0.9376, which proved a more appropriate model fit.

The result in Fig. 5 shows that there are two groups centred around “Category”, “Review” and “Install” were playing roles of important factors. It shows that “Category” is the most important factor determining “Rating”, and the determinant of Rating in the second internal node differs by category. One is determined by “Review”, while the other by “Install”. Adopting “Review” as the criteria, in the group with a high volume of reviews, “subjectivity” was a determinant for higher rating, the more subjectivity was associated with the high rating. In contrast, in the group with a small number of reviews, it was possible to observe a complex relationship with other variables. Dividing apps into two groups by category, apps with greater polarity are positively associated with a higher rating.

In the group with a low number of installations, a certain category was a determinant of a higher rating. Meanwhile, in the group with higher number of installations, there was a complex relationship between variables: category played a function of the first determinant for a higher rating, followed by review as the second determinant, such that the higher the number of reviews, the higher the rating. Furthermore, the result partly supported the results from the linear regression model, suggesting that in general the greater the number of reviews, the higher the app rating would be, despite the overturn in some categories with a negligible quantity.

Fig. 5. Random Forest Tree

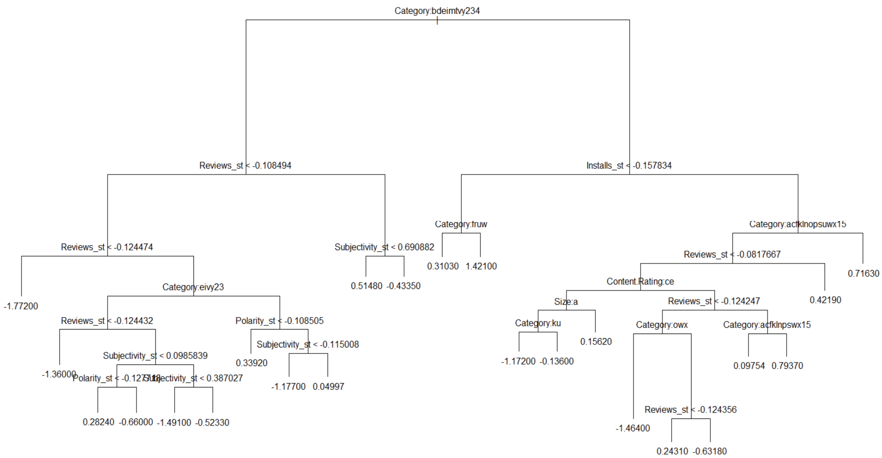
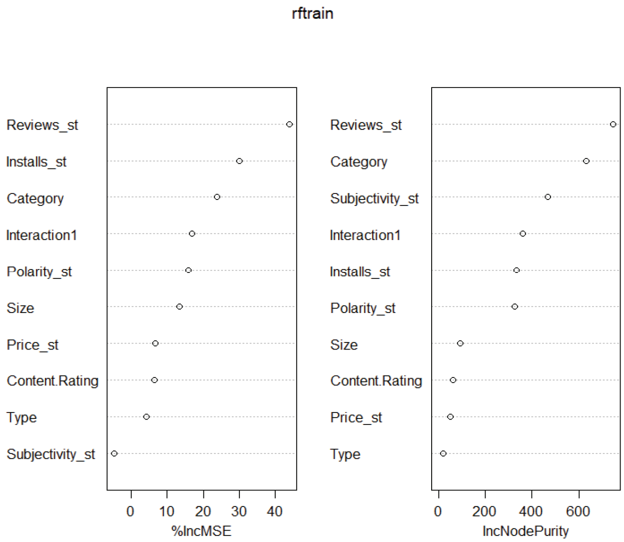


Fig. 6 depicts the results of the investigation into the significance of each variable. Figures in the left-hand graph correspond to measurements of the average reduction in prediction accuracy in the case of ruling out a given variable in a model, while the right graph represents the average reduction of node impurity for all trees derived from the division by a given variable. These results show that “Review” (%lncMSE=44) and “Install” (%lncMSE=30) were the most significant variables in determining “Rating”, followed, in order, by “Category”, the interaction term and “Polarity”.

Fig. 6. Random Forest Tree



4.4. Result of Neural Network Model

4.4.1. Neural Network

It is important to observe to what extent each independent variable influences the dependent variable. The size of the variance of individual independent variables to the dependent variable is a useful tool for doing this, as it is possible to say that a particular variable has an influence on the dependent variable when its variance approaches zero. Since this was not the case for all variables in this experiment, it can be concluded that not all variables influence the dependent variable.

Fig. 7. Generalized weight of every single variable (significance of variables)

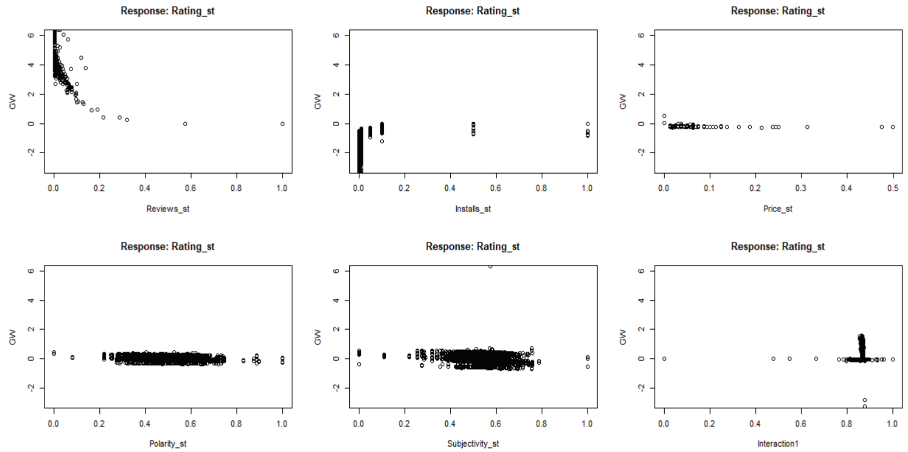
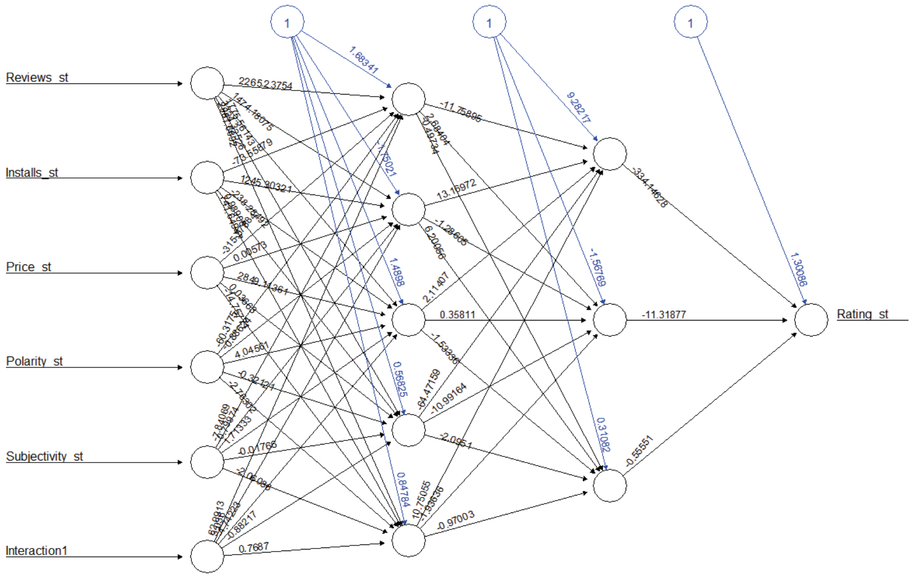


Fig. 8 is the visual result of a neural network model. The black line represents the correlation between and among every single layer, while every single correlation has its own weight. In the neural network model, the influence of each variable can be observed by generalised weight.

Fig. 8. Neural Network Model

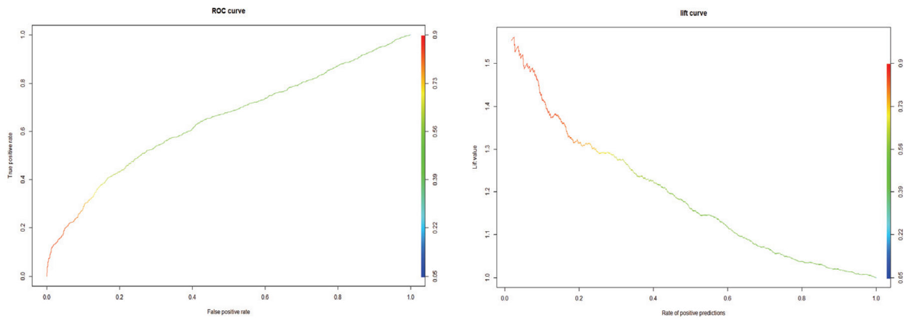


4.4.2. Robustness Test

As for robustness test concerned, credibility, validity and usefulness of a neural network model has been examined by evaluating the Lift and the ROC curves aiming at deciding its prominence. The appropriateness of a prediction model occurs when the lift of the high rating is bigger becoming into be sharply reduced as it moves into the low rating (Berry and Linoff, 1997). Meantime, the ROC curve provided proportional increase in sensitivity and specificity (Berry and Linoff, 1997), which are two criterion for evaluating the efficiency of a model.

Fig. 9. Credibility and Validity of a Neural Network Model

(a) ROC Curve (b) Lift Curve



4.5. Result Summary: Predictors and Factors

Throughout various trials, with the lowest RMSE of 0.1278 among numerous models, the Neural Network model seemed to be the most appropriate choice. Research models can be evaluated in terms of their accuracy and explanatory power. A model may be limited by its lack of flexibility, providing only a narrow range of functions for the estimation of a given function. For instance, a linear regression model is limited by its linearity; the polynomials of a smoothing spline allow for improved function estimation (James et al., 2013, 24). There is no "one size fits all" model; the choice of model depends on the research objective. For instance, a linear regression model is well suited for inference as it has the greatest explanatory power but is limited in its flexibility, while for the purposes of function prediction, greater flexibility may be preferred. Since this paper aims to infer the determinants of an app's rating, a linear regression model was chosen. This study also considers whether additional factors may influence the rating, taking account the significance of these other factors, as deduced from a neural network model.

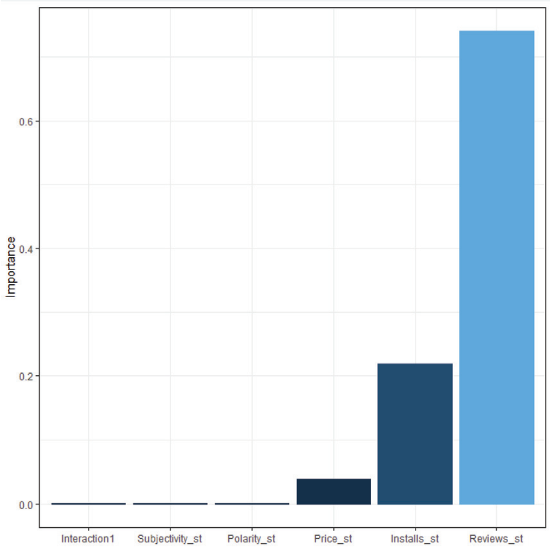
Table 5. Model Comparison

Model	MSE	RMSE
Multilinear Regression	17.7223	4.2097
Regression Tree	1.0121	1.0060
Random Forest Tree	0.8791	0.9376
Neural Network	0.0163	0.1278

Fig. 10 depicts the importance of all the factors in a neural network model. "Review" is the most significant, followed by "Install" and then "Price", while "Polarity", "Subjectivity" and the interaction term are less significant. This result partially contrasts with the results of the abovementioned multiple linear regression model. Nevertheless, it does not necessarily imply that a neural network model with greater predictive ability tells us a different story about the determinants of an app's rating. This is because a linear regression model allows us to understand the causality between predictors and a dependent variable, while a neural network model facilitates the recognition of the main factors. That is, the determinants deduced from a linear regression model are significantly positively or negatively associated with a dependent variable. In contrast, the analytic results of a neural network model, based on a weight-partitioning method, describe how contributing factors influence the probability of a dependent variable (Garson, 1991; Harvey, 1994). A regression model estimates the causality between variables by employing a parametric method, while a neural network model employs either semi-parametric or non-parametric methods. Therefore, only the former allows for an estimation of the extent to which a regression coefficient contributes to the model (Garson, 1991; Harvey, 1994; Kim Myoung-Jong, 2012). Accordingly, researchers point to the complementary advantages of each approach, rather than claiming that either enjoys outright superiority (Kim Tae-Hoon and Hong Han-Kuk, 2004; Lee Kun-Chang, Han In-Goo and Kim Myoung-Jong, 1996; Lee Kun-Chang, Kim Myoung-Jong and Kim Hyuk, 1994).

The results of an explanatory model test, based on a linear regression model, indicate that hypothesis 1 (a positive correlation between the number of times the app has been installed and its rating) is not supported. Hypothesis 2 (a positive causality between review and rating) and hypothesis 3 (that the relationship between polarity and rating would not be significant) are both supported. Hypothesis 4 (that review has a moderating function on the polarity-rating relationship) and hypothesis 5 (no significant relationship between price and rating) are also supported. Finally, hypothesis 6a (no significant relationship between subjectivity and rating) and hypothesis 6b (no significant impact on the reviews-rating relationship) are supported.

Fig. 10. Importance of factors in a neural network model



On the other hand, the result of a prediction model test employing a neural network model proved that review, install and price are the main factors associated with rating, which is partially consistent with the determinants arising from the explanatory model. In particular, review proved to be both a determinant and main factor of an app's rating. Meanwhile, despite not being significant determinants of rating, install and price do seem to have predictive power for an app's rating. Table 6 contains the results of the test of these seven hypotheses using the linear regression model.

Table 6. Result Summary

Hypotheses	Prediction (correlation)	Result
H1	Install -> Rating (+)	Not Supported
H2	Review -> Rating (+)	Supported
H3	Polarity -> Rating (0)	Supported
H4	Review*Polarity -> Rating (+)	Supported
H5	Price -> Rating (0)	Supported
H6a	Subjectivity -> Rating (0)	Supported
H6b	Review*Subjectivity -> Rating (0)	Supported

Note: (+) means positive correlation, while (0) means no correlation.

In summary, from the linear relationship between variables it was found that Review is a predictor that is directly and positively associated with Rating, and that it also has a moderating effect on the Polarity-Rating relationship. On the other hand, from a nonlinear perspective, it was found that Review, Install and Price have a significant influence on Rating.

5. Discussion

5.1. Theoretical Implications

Our results represent an extension to the extant literature in two key areas: (1) the mediator of emotions between motivations and outcomes, and (2) the moderating effect of a behavioral factor on the motion-outcome relationship. Firstly, from the results, we gain a general understanding of the role of emotions observable from consumer behavior as a mediator. Fisher and Fisher (2002)'s information-motivation-behavioral skills model assumes that information and motivation can predict people's behavioral skills, and that behavioral skills affect people's ultimate behavior. This model therefore suggests that information, motivation and behavioral skills are the three most influential factors that cause behavioral change in people. However, this model does not consider the cause-and-effect relationship between motivation and behavior, which remains a black box. These gaps were filled by Dawson, Bloch and Ridgway (1990)'s motive-emotion-outcome model, which suggests that people's emotions are mediators in the link between people's purchasing motives and their purchasing behaviors (Dawson, Bloch and Ridgway, 1990). Such purchasing motivations include product and experimental motives (Dawson, Bloch and Ridgway, 1990), hedonic shopping motives, utilitarian shopping motives (Babin, Darden and Griffin, 1994) and product-oriented and empirical motives (Westbrook and Black, 1985), as well as preferences (Dawson, Bloch and Ridgway, 1990), attitudes toward shopping and brand image and willingness to patronize, or spend more time in a particular store (Donovan and Rossiter, 1982).

Secondly, theoretical investigations were performed regarding the role of behavioral factors, such as user-written reviews, in outcomes based on people's emotions and perceptions. Emotions, which are known to be mediators of the cause-and-effect relationship between purchasing motives and behavior, were also found to have been influenced by reviews—a behavioral aspect. This creates the possibility for new theoretical insights in the role of behavior in affecting people's emotions when purchasing a product. According to this research, having more reviews not only directly improves product evaluation scores, but also indirectly shows a proportional, positive relationship with ratings, regardless of the quality of the product itself because of an inability of the emotional and perceptional variable to cause significant changes in the rating. This finding provides a complementary explanation to the black box in the linear relationship put forward by Dawson, Bloch and Ridgway (1990), Liao et al. (2016) and Zeelenberg and Pieters (2004), which explained the relationship between emotional and perceptional factors on behavioral elements.

5.2. Managerial Implications

In this paper, we have found that the number of reviews of an app has a direct positive relationship with its average rating, while there is no positive or negative relationship between emotions and ratings, which was once thought to be statistically insignificant. This has some important implications for global management. Firstly, there is a need for marketers and managers to move away from the traditional, rigid notion that a positive emotional state will automatically cause satisfaction and lead to brand loyalty. Consumers who experienced positive emotions can be divided into two categories – those who have a promotion focus and those who have a prevention focus (Higgins, 1996, 2000). Although both types upload reviews, people with a promotion focus include creative content in their writing, such as new ideas to improve the app, while those with a prevention focus are orientated more towards maintaining the current quality of the service, providing information on a purely narrative and descriptive level (Higgins, 1996). Similarly, consumers with a negative response to an app can also be divided into two camps—passive and active customers. Passive consumers, despite their dissatisfaction, are unlikely to complain directly and so do not contribute information to improve the app's rating. On the other hand, active consumers will write reviews to complain contributing ideas to improve the app.

Secondly, the relationship between review volume and polarity, which are the main attributes of reviews, can be a strategic resource for managers. In this study, there were no Granger relationship between the two variables, nor was there a cause-and-effect relationship. Therefore, the polarity of the review is not a significant aspect for companies to consider when establishing their strategy; rather, they would do well to focus on "trust" and "empathy". For users, reviews serve as easily accessible information with reduced costs (Chen, Wu and Yoon, 2004; Liu, Karahanna and Watson, 2011), serving to reduce the uncertainty around product quality (Li and Hitt, 2010; Zhu and Zhang, 2010) and resolving inconsistencies in the information available (Lizzeri, 1999). Additionally, the excitement and enjoyment of previous users give rise to increased reliability, assurance, responsiveness and empathy for the quality of information conveyed in reviews (Wakefield and Blodgett, 1999). As a result, users will consciously assess the quality of information from these reviews, strengthening the behavioral intentions of further exploring and communicating through the app (Mehrabian and Russell, 1974). Furthermore, users are more likely to trust reviews compared with other product information (Liu, 2006; Sher and Lee, 2009), and this trust is amplified further by user empathy (Xia and Bechwati, 2008). Other users become more willing to write their own reviews in turn (Boulding et al., 1993), which is analogous to the effects of word-of-mouth marketing in offline markets.

6. Conclusion

To sum up, just as management based on customer experience affects emotion (Kidwell, Hardesty and Childers, 2008), reviews manipulate positive and negative emotions to change aspects of information and improve behavioral results. The app market, selling experience goods and digital goods, is expanding rapidly and both enjoys and suffers from unique accessibility features (no trials, no face-to-face contact with sellers, remote sales etc.). In light of this, app developers should prioritize user experience management.

Despite the suggestive findings, this paper does not provide a full explanation of how a user decides what rating to give an app. Indeed, there is a variety of follow-up questions for research stemming from this paper. Firstly, this research covers the role of user reviews in the relationship between positive and negative emotions and ratings beyond the scope of cause-and-effect. Further research could consider whether subjectivity of user reviews directly or indirectly correlates with an app's rating in a great detail. Secondly, user emotions could be treated with greater granularity – rather than a simple binary categorization of emotions as either "positive" or "negative", a whole range might be considered, e.g. regret or disappointment, feelings of exclusion or attachment, depression, pleasure, interest or arousal, dominance, relaxation, contented, satisfied, happiness, astonishment, agitation, rewarded, joyfulness, excitement, pride, as well as negative emotions, such as concern, discomfort, ignorance, anger (Mehrabian and Russell, 1974; Zeelenberg and Pieters, 2004). Thirdly, studies might look into the effect of emotion on purchasing behavior by analyzing the time series data of consumers and comparing the difference between their purchase experience and that of potential buyers. Finally, this paper is not free from methodological errors. Given the impact of demographic attributes on psychological loyalty, satisfaction, repurchase intention, and intention to recommend others (Darley and Smith, 1995; Harrington, Ottenbacher and Kendall, 2011; Putrevu, 2001), a review on the impact of demographic attributes including nationality on app evaluation should be included in future studies to obtain a more complete analysis result.

Studies on the factors that determine app ratings may be more comprehensive if they address the follow-up topics outlined above. Hopefully this research will encourage others to continue research on this topic.

References

Aaker, D. A. (1995), *Building Strong Brands*, New York, NY: Free Press.

Akaike, H. (1973), "Maximum Likelihood Identification of Gaussian Autoregressive Moving Average Models", *Biometrika*, 60(2), 255-265.

Anderson, D. J., T. A. Sweeney and T. A. Williams (1996), *Statistics for Business and Economics*, Minneapolis, MN: St. Paul, West Publishing.

Anderson, E. W. and M. W. Sullivan (1993), "The Antecedents and Consequences of Customer Satisfaction for Firms", *Marketing Science*, 12(2), 125-212.

Babin, B. J., W. R. Darden and M. Griffin (1994), "Work and/or Fun: Measuring Hedonic and Utilitarian Shopping Value", *Journal of Consumer Research*, 20(4), 644-656.

Baier, D. and E. Stüber (2010), "Acceptance of Recommendations to Buy in Online Retailing", *Journal of Retailing and Consumer Services*, 17(3), 173-180.

Baker, D. A. and G. P. Algorta (2016), "The Relationship between Online Social Networking and Depression: A Systematic Review of Quantitative Studies", *Cyberpsychology, Behavior, and Social Networking*, 19(11), 638-648.

Bardus, M., S. B. van Beurden, J. R. Smith and C. Abraham (2016), "A Review and Content Analysis of Engagement, Functionality, Aesthetics, Information Quality, and Change Techniques in The Most Popular Commercial Apps for Weight Management", *International Journal of Behavioral Nutrition and Physical Activity*, 13, 1-9. Available from https://doi.org/10.1186/s12966-016-0359-9

Barron, A. R. (1993), "Universal Approximation Bounds for Super Positions of A Sigmoidal Function", *IEEE Transactions on Information Theory*, 39(3), 930-945.

Berger, J. and E. M. Schwartz (2011), "What Drives Immediate and Ongoing Word of Mouth", *Journal of Marketing Research*, 48(5), 869-880.

Berger, J., A. T. Sorensen and S. J. Rasmussen (2010), "Positive Effects of Negative Publicity: When Negative Reviews Increase Sales", *Marketing Science*, 29(5), 815-827.

Berry, M. and G. Linoff (1997), *Data Mining Techniques: For Marketing, Sales, and Customer Support*, New York, NY: John Wiley & Sons.

Boulding, W., A. Kalra, R. Staelin and V. A. Zeithaml (1993), "A Dynamic Process Model of Service Quality: From Expectations to Behavioral Intentions", *Journal of Marketing Research*, 30(1), 7-27.

Brown, J. J. and P. H. Reingen (1987), "Social Ties and Word-of-Mouth Referral Behavior", *Journal of Consumer Research*, 14(3), 350-362.

Brown, S. R. (1980), *Political Subjectivity*, New Haven, CT: Yale University Press.

Chatterjee, S. and B. Price (1991), *Regression Analysis by Example*, New York, NY: John Wiley & Sons.

Chen, P. Y., S. Wu and J. Yoon (2004), "The Impact of Online Recommendations and Consumer Feedback on Sales", *International Conference on Information Systems 2004 Proceedings*. 711-724. Available from https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1146&context=icis2004

Cheung, C. M. K. and M. K. O. Lee (2008), "The Impact of Electronic Word-of-Mouth: The Adoption of Online Opinions in Online Customer Communities", *Applications and Policy*, 18(3), 229-247.

Chevalier, J. A. and D. Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews", *Journal of Marketing Research"*, 43(3), 345-354.

Childer, T., C. Carr, J. Peck and S. Carson (2001), "Hedonic and Utilitarian Motivations for Online Retail Shopping Behavior", *Journal of Retailing*, 77(4), 511-535.

Cho, Hyu-Kjun, Ju-Young Kang and Dae-Yong Jeong (2016), "An Exploratory Study on Mobile App Review through Comparative Analysis between South Korea and US", *Journal of Information Technology Services*, 15(2), 169-184.

Choi, Jong-Hoo (2000), *Analysis of Data Mining Decision Tree by AnswerTree*, Seoul: SPSS Academy.

Darley, W. K. and R. E. Smith (1995), "Gender Differences in Information Processing Strategies: An Empirical Test of The Selectivity Model in Advertising Response", *Journal of Advertising*, 24(1), 41-56.

Dawson, S., P. Bloch and N. Ridgway (1990), "Shopping Motives, Emotional States, and Retail Outcomes", *Journal of Retailing*, 66(4), 408-427.

Dellarocas, C., N. Awad and M. Zhang (2005), *Using Online Ratings as a Proxy of Word-of-Mouth in Motion Picture Revenue Forecasting* (SSRN Working Paper). Available from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.362.7819&rep=rep1&type=pdf

Diamantopoulos, A. and J. A. Siguaw (2006), "Formative versus Reflective Indicators in Organizational Measure Development: A Comparison and Empirical Illustration", *British Journal of Management*, 17(4), 263-282.

Donovan, R. J. and J. R. Rossiter (1982), "Store Atmosphere: An Environmental Psychology Approach", *Journal of Retailing*, 58(1), 34-57.

Filieri, R. (2016), "What Makes An Online Consumer Review Trustworthy?", *Annals of Tourism Research*, 58(May), 46-64.

Fisher, J. D. and W. A. Fisher (2002), "The Information-Motivation-Behavioral Skills Model". In R. J. Diclemente, R. A. Crosby and M. C. Kegler (Eds.), *Emerging Theories in Health Promotion Practice and Research*, New York, NY: John Wiley & Sons, 40-70.

Franses, P. H. and K. Van Griensven (1998), "Forecasting Exchange Rates Using Neural Networks for Technical Trading Rules", *Studies in Nonlinear Dynamics & Econometrics*, 2(4), 109-114.

Frechtling, D. (2001), *Forecasting Tourism Demand Methods and Strategies*, Oxford, UK: Butterworth-Heinemann.

Frie, K., J. Hartmann-Boyce, S. Jebb, C. Albury, R. Nourse and R. Aveyard (2017), "Insights From Google Play Store User Reviews for the Development of Weight Loss Apps: Mixed-Method Analysis", *JMIR Mhealth Uhealth*, 5(12), 1-14.

Garson, D. G. (1991), "Interpreting Neural Network Connection Weights", *AI Expert*, 6(7), 47-51.

Gefen, D., V. S. Rao and N. Tractinsky (2003), "The Conceptualization of Trust, Risk and Their Relationship in Electronic Commerce: The Need for Clarifications", *Proceedings of the 36th Hawaii International Conference on System Sciences*.

Griffiths, W. E., R. C. Hill and G. G. Judge (1993), *Learning and Practicing Econometrics*, New York, NY: John Wiley & Sons.

Gu, J., Y. C. Xu, H. Xu, C. Zhang and H. Ling (2017), "Privacy Concerns for Mobile App Download: An Elaboration Likelihood Model Perspective", *Decision Support Systems*, 94(February), 19-28.

Gupta, P. and J. Harris (2010), "How e-WOM Recommendations Influence Product Consideration and Quality of Choice: A Motivation to Process Information Perspective", *Journal of Business Research*, 63(9-10), 1041-1049.

Hair, J. F. J., R. E. Anderson, R. L. Tatham and W. C. Black (1998), *Multivariate Data Analysis* (5th ed.), Englewood Cliffs, NJ: Prentice-Hall.

Hammond, K., G. McWIlliam and A. N. Diaz (1998), "Fun and Work on the Web: Differences in Attitudes between Novices and Experienced User". In J. W. Alba and J. W. Hutchinson (Eds.), *NA-Advances in Consumer Research* (Vol. 25), Provo, UT: Association for Consumer Research, 372-378.

Harman, M., Y. Jia and Y. Zhang (2012), "App Store Mining and Analysis: MSR for App Stores", 2012 9th IEEE Working Conference on Mining Software Repositories (MSR). Available from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6224306

Harrington, R., M. Ottenbacher and K. Kendall (2011), "Fine Dining Restaurant Selection: Direct and Moderating Effects of Customer Attributes", *Journal of Foodservice Business Research*, 14(3), 272-289.

Harvey, R. L. (1994), *Neural Network Principles*, Englewood Cliffs, NJ: Prentice-Hall.

Hausman, A. V. and J. S. Siekpe (2009), "The Effect of Web Interface Features on Consumer Online Purchase Intentions", *Journal of Business Research*, 62(1), 5-13.

Helsel, D. R. and R. M. Hirsch (1992/2002), *Statistical Methods in Water Resources*, Amsterdam: Elsevier.

Herr, P. M., F. R. Kardes and J. Kim (1991), "Effects of Word-of-Mouth and Product-Attribute Information on Persuasion: An Accessibility-Diagnosticity Perspective", *Journal of Consumer Research*, 17(4), 454-462.

Higgins, J. R. (1996/2000), *Sampling Theory in Fourier and Signal Analysis: Foundations*, Oxford, UK: Clarendon Press.

Hoffman, D. L. and T. P. Novak (1996), "Marketing in Hypermedia Computer-mediated Environments: Conceptual Foundations", *Journal of Marketing*, 60(3), 50-68.

Hoon, L., R. Vasa and J. G. Schneider (2013), *An Analysis of the Mobile App Review Landscape: Trends and Implications* (Unpublished Paper), Melbourne, Australia: Swinburne University of Technology. Available from https://researchbank.swinburne.edu.au/file/ba489b3c-a9bb-41f0-8b43-b6d1ff4d4829/1/PDF%20(23%20pages).pdf

Lacob, C. and R. Harrison (2016), "Retrieving and Analyzing Mobile Apps Feature Requests from Online Reviews", *2013 10th Working Conference on International Mining Software Repositories*, 41-44. Available from https://ieeexplore.ieee.org/xpl/conhome/6597024/proceeding

James, G., D. Witten, T. Hasite and R. Tibshirani (2013), *An Introduction to Statistical Learning: With Applications in R*, New York, NY: Springer.

Keller, K. L. (1993), "Conceptualizing, Measuring, and Managing Customer-Based Brand Equity", *Journal of Marketing*, 57(1), 1-22.

Kennedy, P. (1992), *A Guide to Econometrics*, Cambridge, MA: The MIT Press.

Kidwell, B., D. M. Hardesty and T. L. Childers (2008), "Consumer Emotional Intelligence: Conceptualization, Measurement, and The Prediction of Consumer Decision Making", *Journal of Consumer Research*, 35(1), 154-166.

Kim, D. J., D. L. Ferrin and H. R. Rao (2008), "A Trust-based Consumer Decision-making Model in Electronic Commerce: The Role of Trust, Perceived Risk, and Their Antecedents", *Decision Support Systems*, 44(1), 544-564.

Kim, Gi-Mun and Hoon-Young Koo (2016), "The Causal Relationship between Risk and Trust in The Online Marketplace: A Bidirectional Perspective", *Computers in Human Behavior*, 55(February), 1020-1029.

Kim, Myoung-Jong (2012), "Performance Comparison of Internal Accounting Control Assessment Models Applying Logistic Regression and Neural Networks", *Korean International Accounting Review*, 46, 1-30.

Kim, Sang-Hwan (2000), "Establishing an Optimal Neural Network Model and Analyzing The

Performance of Foreign Exchange Prediction", *Financial Research*, 14(1), 57-85.
Kim, Tae-Hoon and Han-Kuk Hong (2004), "A Study on Apartment Price Models Using Regression Model and Neural Network Model", *The Korea Spatial Planning Review*, 43, 183-200.
Kuan, C. M. and T. Liu (1995), "Forecasting Exchange Rates Using Feedforward and Recurrent Neural Networks", *Journal of Applied Econometrics*, 10(4), 347-364.
Lee, Sang-Jae and Joon-Yeon Choeh (2014), "Predicting The Helpfulness of Online Reviews Using Multilayer Perceptron Neural Networks", *Expert Systems with Applications*, 41(6), 3041-3046.
Lee, Dong-Il and Seung-Hoon Choi (2012), "The Impact of Consumer Review and Expert Review on the App Developer's Performance in The App Store", *Journal of Korean Marketing Association*, 27(2), 113-136.
Lee, Eun-Ju and Soo-Yun Shin (2014), "When Do Consumers Buy Online Product Reviews? Effects of Review Quality, Product Type, and Reviewer's Photo", *Computers in Human Behavior*, 31(February), 356-366.
Lee, Kook-Yong (2017), "The Effects of E-WOM in Selecting the Mobile Application", *The Journal of the Korea Contents Association*, 17(1), 80-91.
Lee, Kun- Chang, In-Goo Han and Myoung-Jong Kim (1996), "A Study on The Credit Evaluation Model Integrating Statistical Model and Artificial Intelligence Model", *Journal of Management Science*, 21(1), 81-100.
Lee, Kun-Chang, Myoung-Jong Kim and Hyuk Kim (1994), "An Inductive Learning-Assisted Neural Network Approach to Bankruptcy Prediction: Comparison with MDA, Inductive Learning, and Neural Network Models", *Journal of Management Research*, 23(3), 109-144.
Leonard-Barton, D. (1985), "Experts as Negative Opinion Leaders in The Diffusion of a Technological Innovation", *Journal of Consumer Research*, 11(4), 914-926.
Li, X. and L. M. Hitt (2010), "Price Effects in Online Product Reviews: An Analytical Model and Empirical Analysis", *MIS Quarterly*, 34(4), 809-831.
Liao, C., H. N. Lin, M. M. Luo and S. Chea (2016), "Factors Influencing Online Shoppers' Repurchase Intentions: The Roles of Satisfaction and Regret", *Information & Management*, 54(5), 651-668.
Liu, B. (2010), "Sentiment Analysis and Subjectivity". In N. Indurkaya and F. J. Damerau (Eds.), *Handbook of Natural Language Processing* (2nd ed.), Cambridge, UK: Chapman & Hall Book, 627-665.
Liu, Q. B., E. Karahanna and R. T. Watson (2011), "Unveiling User-generated Content: Designing Websites to Best Present Customer Reviews", *Business Horizons*, 54(3), 231-240.
Liu, Y. (2006), "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue", *Journal of Marketing*, 70(3), 74-89.
Lizzeri, A. (1999), "Information Revelation and Certification Intermediaries", *The RAND Journal of Economics*, 30(2), 214-231.
Lovett, M. J., R. Peres and R. Shachar (2013), "On Brands and Word of Mouth", *Journal of Marketing*, 50(4), 427-444.
Martens, D. and T. Johann (2017), "On the Emotion of Users in App Reviews", *IEEE/ACM 2nd International Workshop on Emotion Awareness in Software Engineering*. Available from https://ieeexplore.ieee.org/document/7961885
Mehrabian, A. and J. A. Russell (1974), *An Approach to Environmental Psychology*, New York, NY: MIT Press.
Moody, J. E. (1994), "Prediction Risk and Architecture Selection for Neural Networks". In V. Cherkassky, J. H. Friedman and H. Wechsler (Eds.), *From Statistics to Neural Networks, NATO ASI Series (Series F: Computer and Systems Sciences, vol. 136),* Berlin: Springer, 147-165. Available from https://doi.org/10.1007/978-3-642-79119-2_7
Nayebi, Maleknaz, H. Cho and G. Ruhe (2018), "App Store Mining Is Not Enough for App

Improvement", *Empirical Software Engineering*, 23(1), 2764-2794.

Oliver, R. L. (1993), "A Conceptual Model of Service Quality and Service Satisfaction: Compatible Goals, Different Concepts", *Advances in Services Marketing and Management*, 2(1), 65-85.

Oliver, R. L. (1997), *Satisfaction: A Behavioral Perspective on The Consumer*, New York, NY: McGraw-Hill.

Oliver, R. L. and W. S. DeSarbo (1988), "Response Determinants in Satisfaction Judgments", *Journal of Consumer Research*, 14(4), 495-507.

Pagano, D. and W. Maaleg (2013), "User Feedback in The Appstore: An Empirical Study", *2013 21st IEEE International Conference on Requirements Engineering*. Available from https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6636712

Palomba, F. and M. Linares-Vasquez (2015), "User Reviews Matter! Tracking Crowdsourced Reviews to Support Evolution of Successful Apps", *IEEE International Conference on Software Maintenance and Evolution*. Available from https://ieeexplore.ieee.org/document/7332475

Parasuraman, A., V. A. Zeithaml and L. L. Berry (1991), "Understanding Customer Expectations of Service", *Sloan Management Review*, 32(Spring), 39-48.

Park, C. W., B. J. Jaworski and D. J. Maclnnis (1986), "Strategic Brand Concept-Image Management", *Journal of Marketing*, 50(4), 135-145.

Pavlou, P. A. (2003), "Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with The Technology Acceptance Model", *International Journal of Electronic Commerce*, 7(3), 101-134.

Pavlou, P. A. and D. Gefen (2004), "Building Effective Online Marketplaces with Institution-based Trust", *Information Systems Research*, 15(1), 37-59.

Petter, S., D. Straub and A. Rai (2007), "Specifying Formative Constructs in IS Research", *MIS Quarterly*, 31(4), 623-656.

Phillips, D. M. (1999), *The Role of Consumption Emotions in the Satisfaction Response*, (Doctoral Dissertation), Philadelphia, PA: Pennsylvania State University.

Phillips, P., K. Zigan, M. M. S. Silva and R. Schegg (2015), "The Interactive Effects of Online Reviews on The Determinants of Swiss Hotel Performance: A Neural Network Analysis", *Tourism Management*, 50(1), 130-141.

Ponte, E. B., E. Carvajal-Trujillo and T. Escobar-Rodriguez (2015), "Influence of Trust and Perceived Value on the Intention to Purchase Travel Online: Integrating The Effects of Assurance on Trust Antecedents", *Tourism Management*, 47(April), 286-302.

Putrevu, S. (2001), "Exploring The Origins and Information Processing Differences between Men and Women: Implications for Advertisers", *Academy of Marketing Science Review*, 10(1), 1-14.

Ripley, B. D. (1993), "Statistical Aspects of Neural Networks". In O. E. Barndorff-Nielsen, J. L. Jensen and W. S. Kendall (Eds.), *Networks and Chaos – Statistical and Probabilistic Aspects*, London, UK: Chapman & Hall, 40-111.

Schmitt, P., B. Skiera and C. V. den Bulte (2011), "Referral Programs and Customer Value", *Journal of Marketing*, 75(1), 46-59.

Sher, P. J. and S. H. Lee (2009), "Consumer Skepticism and Online Reviews: An Elaboration Likelihood Model Perspective", *Social Behavior and Personality: an International Journal*, 37(1), 137-143.

Swanson, N. R. and H. White (1995), "A Model-selection Approach to Assessing The Information in The Term Structure Using Linear Models and Artificial Neural Networks", *Journal of Business & Economic Statistics*, 13(3), 265-275.

Swinyard, W. R. (1993), "The Effects of Mood Involvement and Quality of Store Experience on Shopping Intentions", *Journal of Consumer Research*, 20(2), 271-280.

Tauber, E. M. (1972), "Marketing Notes and Communications: Why Do People Shop?", *Journal of Marketing*, 36(4), 46-59.

Wakefield, K. L. and J. G. Blodgett (1999), "Customer Response to Intangible and Tangible Service Factors", *Psychology & Marketing*, 16(1), 51-68.

Weiss, S. M. and C. A. Kulikowski (1991), *Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems*, San Francisco, CA: Morgan Kaufmann Publishers Inc.

Westbrook, R. A. and W. C. Black (1985), "A Motivation-based Shopper Typology", *Journal of Retailing*, 61(1), 78-103.

Witt, S. F. and C. A. Witt (1995), "Forecasting Tourism Demand: A Review of Empirical Research", *International Journal of Forecasting*, 11(3), 447-475.

Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT press.

Xia, L. and N. N. Bechwati (2008), "Word of Mouse: the Role of Cognitive Personalization in Online Consumer Reviews", *Journal of Interactive Advertising*, 9(1), 3-13.

Yoo, Chang-Jo, Jong-Hee Park and D. Maclnnis (1998), "Effects of Store Characteristics and In-Store Emotional Experiences on Store Attitude", *Journal of Business Research*, 42(3), 253-263.

Yuan, L. (2015), "Kingmakers of China's Internet: Baidu, Alibaba and Tencent", *The Wall Street Journal*. Available from http://www.wsj.com/articles/kingmakers-of-chinas-internet-baidu-alibaba-and-tencent-1445451143?mod=ST

Zeelenberg, M. and R. Pieters (2004), "Beyond Polarity in Customer Dissatisfaction: A Review and New Findings on Behavioral Responses to Regret and Disappointment in Failed Services", *Journal of Business Research*, 57(4), 445-455.

Zhang, W. and S. Watts (2008), "Online Communities as Communities of Practice: A Case Study", *Journal of Knowledge Management*, 12(4), 55-71.

Zhong, N. and F. Michahelles (2013), "Google Play Is Not A Long Tail Market: An Empirical Analysis of App Adoption on the Google Play App Market", *Proceedings of the 28th annual ACM Symposium on Applied Computing*, 499-504.

Zhu, F. and X. Zhang (2010), "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics", *Journal of Marketing*, 74(2), 133-148.