Journal of Organizational and End User Computing
Volume 34 • Issue 6

Predicting User Satisfaction of Mobile Healthcare Services Using Machine Learning:
Confronting the COVID-19 Pandemic

Haein Lee, Department of Applied Artificial Intelligence, Sungkyunkwan University, South Korea
https://orcid.org/0000-0002-9718-6395

Seon Hong Lee, Department of Applied Artificial Intelligence, Sungkyunkwan University, South Korea
https://orcid.org/0000-0003-1112-7974

Dongyan Nan, Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University, South Korea*
Jang Hyun Kim, Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University, South Korea*
https://orcid.org/0000-0001-7750-2664

ABSTRACT

The outbreak of COVID-19 led to rapid development of the mobile healthcare services. Given that user satisfaction is of great significance in inducing marketing success in competition markets, this research explores and predicts user satisfaction with mobile healthcare services. Specifically, the current research aimed to design a machine learning model that predicts user satisfaction with healthcare services using big data from Google Play Store reviews and satisfaction ratings. By dealing with the sentimental features in online reviews with five classifiers, the authors find that logistic regression with term frequency-inverse document frequency (TF-IDF) and XGBoost with bag of words (BoW) have superior performances in predicting user satisfaction for healthcare services. Based on these results, the authors conclude that such user-generated texts as online reviews can be used to predict user satisfaction, and logistic regression with TF-IDF and XGBoost with BoW can be prioritized for developing online review analysis platforms for healthcare service providers.

KEYWORDS

Big Data, BoW, Logistic Regression, Machine Learning, Mobile Healthcare Service, Natural Language Processing, Online Review, TF-IDF, User Satisfaction, XGBoost

INTRODUCTION

Along with the global spread of COVID-19, people's demands for a healthy lifestyle have enhanced rapidly since 2019. Due to the recommendations of the World Health Organization (WHO) (Jones et al., 2020) on maintaining physical distance, healthcare-related services cannot provide face-to-face services. Countering this situation, several companies have been striving to provide mobile

DOI: 10.4018/JOEUC.300766
*Corresponding Author

technology-based healthcare services (i.e., mobile healthcare service) to individuals who desire to take care of their health. Because mobile healthcare services do not require face-to-face communication, the services are suitable for the Covid-19 environment and have attracted a large number of users (Ming et al., 2020).

Mobile healthcare service is conceptualized as a system that offers health-related information and services via mobile communication technologies (e.g., smart phones and mobile networks) (Jaiswal and Anand, 2021; Zhao et al., 2018). Owing to the immense contribution to big data technology and artificial intelligence in recent times, mobile healthcare services can adequately trace the health condition of their users, and offer timely treatment and care (Zhao et al., 2018; Bol et al., 2018; Saheb, 2020). Mobile healthcare services not only save diagnosis time and cost but also play a positive role to improve the efficiency of medical resources (Lin et al., 2021). Therefore, a number of countries (e.g., Korea, China, and United States) have been taking efforts to support the development of mobile healthcare services (Zhao et al., 2018). Additionally, many people have been attracted by the advantages of mobile healthcare services and have resorted in their daily usage (Ming et al., 2020). As of 2021, more than 400,000 mobile healthcare applications have been made available on the Google play store (Imaginovation, 2021). According to Global market insights (2021), the global mobile health market has been estimated to reach $289.4 billion by 2025.

Considering user satisfaction as being related to user loyalty and the intention to use specific technologies continuously (Bhattacherjee, 2001; Kumar et al., 2013; Cheng and Jiang, 2020), several studies have explored the user experience elements that lead to greater user satisfaction with mobile healthcare services (e.g., Oppong et al., 2021; Handayani et al., 2018; Keikhosrokiani et al., 2020). However, most studies exploring the user experience of mobile healthcare services only employed a small number of samples and traditional methodologies (e.g., analyzing small survey-based data with structural equation model). Therefore, to solve these shortcomings, this research attempts to predict and explore user satisfaction with mobile healthcare services by examining big data using machine learning approaches.

Specifically, 139,604 usable online reviews of a particular mobile healthcare service (i.e., Samsung Health) from Google Play store were collected. Subsequently, five machine learning classifiers (i.e., Logistic regression, Random Forest, Gradient Boosting Model, Extreme Gradient Boosting, Naïve Bayes) with three word embedding methods (Bag-of-words (BoW), Term Frequency - Inverse Document Frequency (TF-IDF), Global Vectors for Word Representation (GloVe)) for predicting user satisfaction with healthcare services were applied. It is worth noting that, this is one of the first studies to predict user satisfaction by employing big data and machine learning approaches in the mobile healthcare service field. Theoretically, this study also can contribute to expanding the literature of affect theory (Kratzwald et al., 2018) which indicates that user-generated contents are notably related to user satisfaction with certain services and products.

In order to achieve marketing success in the mobile healthcare service market, which has been fierce due to the spread of COVID-19, prior studies (Alam et al., 2021; Jaana and Paré, 2020; Khalil et al., 2020; Alanzi et al., 2021; Lee and Figueredo, 2021) have implied that it is necessary to investigate and predict the user experience and satisfaction with the services. Therefore, the current research can be regarded as one of the studies contributing to face the COVID-19 pandemic.

LITERATURE REVIEW

User Satisfaction With Mobile Healthcare Services

With reference to the rapid development of mobile healthcare services, many scholars have tried to explore improvements in user experience to enhance user satisfaction. Oppong et al. (2021) employed structural equation modeling (SEM) for computing 305 samples and established that, interaction quality positively influences user satisfaction with mobile healthcare services. Zhang et al. (2018)

used SEM to analyze 273 samples and reported that perceived e-health literacy negatively affects user satisfaction with healthcare services. By computing 127 validated responses with SEM, Handayani et al. (2018) confirmed that ease of access and convenience markedly affect user satisfaction with mobile healthcare systems. Keikhosrokiani et al. (2020) applied covariance-based SEM for computing 323 samples and found that system quality and mHealth literacy influenced user satisfaction with mobile health. Akter et al. (2013a) examined 283 validated data with SEM and concluded that service quality predicts user satisfaction with mobile health services. Based on the SEM outcomes of 350 responses, Akter et al. (2013b) discovered that usefulness and confirmation can lead to a greater user satisfaction in the context of healthcare services. Jaana and Paré (2020) demonstrated that both usefulness and expectation confirmation lead to greater user satisfaction with mobile healthcare technologies by examining the 4,109 samples with SEM approach. Ju and Zhang (2020) reported that service quality positively influences the user satisfaction with healthcare services via the mediation of expectation confirmation by computing 521 survey-based data with SEM.

In reviewing the aforementioned literature on mobile healthcare services, it was found that most of the works explored user satisfaction based on a limited number of samples and traditional approaches (e.g., SEM). To address these limitations, this study explored user satisfaction with mobile healthcare services by using machine learning models to compute large scale online reviews, as suggested by Kim et al. (2021) and Hwang et al. (2020).

User Experience, Online Review, and Machine Learning

Zhao et al. (2019) identified that, exploring online reviews is helpful for understanding user assessments of certain services or products, which is likely to induce marketing success in competitive markets (Yang et al., 2009). Moreover, Chatterjee et al. (2021) suggested that computing user-generated texts such as online reviews is a practical method for detecting user experience with specific services. This is because online reviews not only contain users' opinions and emotions but are also easy to access (Zhao et al., 2019; Chatterjee et al., 2021; Kim et al., 2021). Furthermore, Kim et al. (2021), Wang and Goh (2020), and Jang and Yi (2017) demonstrated that online reviews contain user experience-related elements that may be associated to user satisfaction with particular services and products. Therefore, it is appropriate to predict user satisfaction by analyzing online reviews.

In line with the above-mentioned viewpoints, Zheng et al. (2021), Schaeffer and Sanchez (2020) suggested machine learning approaches (e.g., logistic regression, random forest) as appropriate for exploring user experiences. Furthermore, several scholars have attempted to predict user intention or satisfaction by analyzing online reviews with machine learning approaches. Bansal and Srivastava (2018) predicted user satisfaction by computing online reviews with several machine learning algorithms (i.e., support vector machine, Naïve Bayes, logistic regression, random forest) and confirmed a higher accuracy with random forest (accuracy: 90.21%) and Logistic regression (accuracy: 89.72%) compared with other algorithms. Moreover, Zahoor et al. (2020) found that random forest (accuracy: 94%), Naïve Bayes (accuracy: 91%), and logistic regression (accuracy: 91%) have a great accuracy in predicting user satisfaction with restaurant services. Additionally, Shah et al. (2021) confirmed that random forest (accuracy: 93.17%) and logistic regression (accuracy: 90.88%) has a higher accuracy than multinomial and Bernoulli Naïve Bayes models in forecasting user satisfaction with certain products. La et al. (2021) predicted tourists' satisfaction by analyzing their online comments with Logistic regression and found the classifier to have an accuracy of 80.95%. Luo et al. (2021) analyzed online reviews using random forest to predict positive/negative customer ratings for restaurant services. These studies indicate that user satisfaction can be predicted by employing random forest, logistic regression, and Naïve Bayes for classifying online reviews.

For better classification accuracy, several advanced machine learning methodologies (e.g., extreme gradient boosting (XGBoost), gradient boosting model (GBM)) have been employed for detecting user satisfaction with specific services. By adopting several machine learning classifiers (e.g., random forest, logistic regression, XGBoost), Hwang et al. (2020) confirmed XGBoost to have the best accuracy (83.42%) in terms of forecasting users' positive and negative assessments of airline services. Chatterjee et al. (2021) employed XGBoost for computing product reviews and attained best outcomes with forecasting user satisfaction among other classifiers (e.g., random forest). In addition, Xu et al. (2022) applied XGBoost for analyzing users' comments towards doctors and achieved advanced results in predicting patients' satisfaction. Hew et al. (2020) revealed that examining online reviews with GBM has a great power (F1 score: 0.8375) in forecasting user satisfaction with online learnings. Therefore, this study attempted to employ and compare the performance of logistic regression, random forest, Naïve Bayes, XGBoost, and GBM for detecting user satisfaction with mobile healthcare services.

METHODS

Data Collection

Given "Samsung Health" is a well-known mobile healthcare service with approximately 200 million users as of 2020 (Samsung newsroom, 2020; Yonhap News, 2021), it was selected as the target application with 158,222 online reviews collected from Google play store. This dataset consisted of three metadata (i.e., textual reviews, star ratings, and dates) from April 8, 2015 to September 2, 2021.

Preprocessing

User satisfaction, a psychological variable, is conceptualized as users' global assessments of using specific services or products (Kim et al., 2021; Bhattacherjee, 2001). As indicated in affect theory (Kratzwald et al., 2018) and user experience related studies (Li and Zhang, 2014; Ho-Dac et al., 2013; Hu et al., 2019), user satisfaction can be measured by star ratings of online reviews. As the current research focused on predicting user satisfaction, star ratings were selected as a criterion to build sentiment classifiers based on machine learning methods. Specifically, reviews greater than a 3-star rating was regarded as positive, and those less than a 3-star rating as negative. Thus, a negative review indicates user dissatisfaction, while a positive review demonstrates user satisfaction. Furthermore, following Li and Zhang (2014) and Hu et al. (2019), reviews with 3-star ratings were excluded from the dataset.

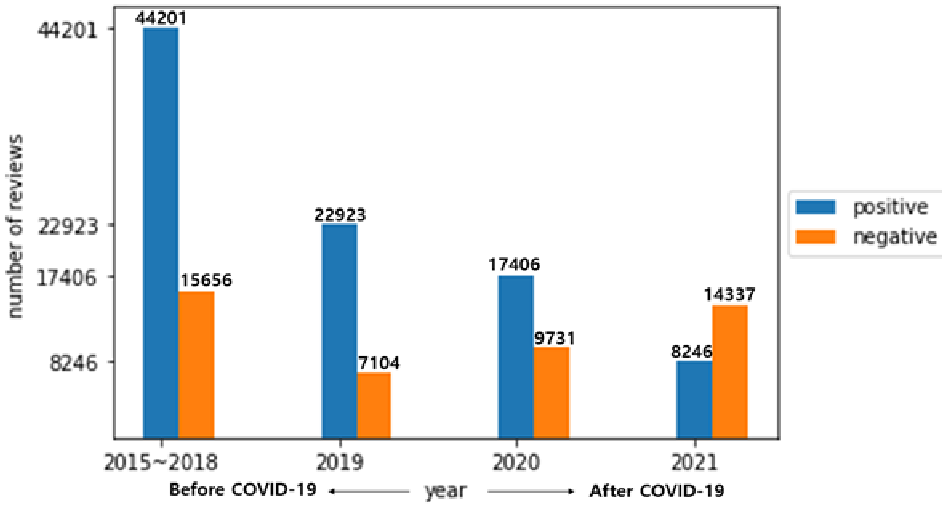
As the COVID-19 pandemic began to spread in January 2020, the authors sliced the dataset into pre-2020 and post-2020 (see Figure 1). In "before COVID-19" dataset, training set consists of data from 2015 to 2018, and the test set consists of data from 2019. In "after COVID-19" dataset, training set consists of data from 2020, and the test set consists of data from 2021.

Before training the classifiers, all textual reviews were preprocessed. In this process, Regular expressions in the Python library were used to remove special symbols. Subsequently, Natural Language Toolkit (NLTK) was used to extract accurate lemmatization and remove the stopwords in NLTK. Consequently, 139,604 reviews were used for the analysis (Figure 1).

Imbalanced Class

For imbalanced datasets, the majority and minority class differences are so large, which may lead the accuracy of minority class prediction to be poor (Chen et al., 2004; Karimi et al., 2019). Thus, owing to the imbalanced distribution of the ratings in the current research, downsampling was performed as suggested by Karimi et al. (2019) and Liu et al. (2021). Specifically, the "before COVID-19" training set and test set contain 59,857 data (positive: 44,201 and negative: 15,656) and 30,027 (positive: 22,923 and negative: 7,104), respectively. Moreover, the "after COVID-19" training set and test set contain 27,137 data (positive: 17,406 and negative: 9,731) and 22,583 data (positive: 8,246 and negative: 14,337), respectively. As a result, the number of positive data was matched with the number of negative data. Overall, "before COVID-19" 31,312 training and 14,208 test datasets and "after COVID-19" 19,462 training and 16,492 test datasets were preprocessed, respectively.

Figure 1. Distribution of positive and negative reviews by year



Text Feature Extraction

Representative methodologies such as BoW, TF-IDF (Joachims, 1997), GloVe (Pennington et al., 2014) were employed for extracting the text features from the data as suggested by Krouska et al. (2020) and Soumya and Pramod (2020).

BoW represents text data in numerical expressions (e.g., word occurrence frequency). These features were obtained by applying the count vectorizer module in scikit-learn, which is a Python library.

TF-IDF indicates how significant a word is in a particular document when a document group includes multiple documents (Joachims, 1997). It computes the frequency by multiplying the term frequency (TF), a value that reveals the frequency of a specific word in a document, and the inverse document frequency (IDF), the relative frequencies corresponding to a set of documents.

The vector W in terms of a document(d) is determined by a combination of the statistics TF (t, d) and DF(t). The term frequency (TF) is word count (t) that appears in document (d). The document frequency (DF) is the number of documents in which the specific word (t) appears. The inverse document frequency $IDF\left(t\right)$ is obtained by document frequency:

$$IDF\left(t\right) = \log\left(\frac{\left|D\right|}{DF\left(t\right)}\right) \tag{1}$$

$\left|D\right|$ is the entire count of documents. The inverse document frequency about a word (IDF) has high value when it arises in a small number of documents. In addition, it has low value when it arises in numerous documents. The certain word feature W is calculated as a dot product (Jing et al., 2002):

$$W = TF\left(t,d\right) * IDF\left(t\right) \tag{2}$$

TF-IDF features were obtained by employing the Tfidf Vectorizer module in scikit-learn.

GloVe is a feature representation method using both count-based and prediction-based. GloVe conducts a dot product of the embedded central word and the surrounding word vector to determine the probability of co-occurrence in the entire corpus (Sakketou and Ampazis, 2020). The GloVe feature representation is easy to use for training and it finds the semantic connections of words in the vector space (Nandanwar and Choudhary, 2021).

The GloVe objective function is made up of Eqn. (3). Objective function $\left(J\right)$ was defined to avoid weighting all co-occurrences equally. X_{ij} means that the number of word (i) appears in the context of word (j). V is the size of the vocabulary. w_i vector and w_j vector are about the main word and the context word respectively. b_i, b_j are scalar biases in terms of the main and context words:

$$J = \sum_{i,j=1}^{V} f\left(X_{ij}\right)\left(w_i^T w_j + b_i + b_j - \log X_{ij}\right)^2 \tag{3}$$

Finally, the weight function $f\left(X_{ij}\right)$ which aids us to avoid learning only from extremely frequent word pairs is as follows (Pennington et al., 2014):

$$f\left(X_{ij}\right) = \begin{cases} \left(\dfrac{X_{ij}}{x_{\max}}\right)^{\alpha} & if\ X_{ij} < x_{\max} \\ 1 & otherwise \end{cases} \tag{4}$$

According to Pennington et al. (2014), $f\left(x\right)$ has several properties. First, $f\left(0\right) = 0$. Second, $f\left(x\right)$ needs to be increasing function in order to prevent overweighting of rare co-occurrences. Third, $f\left(x\right)$ needs to be relatively small for big $X_{ij}(< x_{max})$ in other to block overweighting frequent co-occurrences.

The GloVe model was trained by utilizing its package to find the appropriate frequency values.

Machine Learning Method for Classification

Machine learning models such as logistic regression, random forest, Naïve Bayes, GBM, and XGBoost were employed for classifying positive and negative reviews. Logistic regression, random forest, Naïve Bayes, and GBM were implemented using scikit-learn. However, for random forest and GBM, the ensemble module was employed (Hakak et al., 2021; Konstantinov and Utkin, 2021). Furthermore, the XGBoost model was implemented by utilizing XGBoost python library. At the same time, a grid search, which is provided in scikit-learn, was used for tuning optimized hyperparameters (Budholiya et al., 2020). The process used to classify reviews is shown in Figure 2.

RESULTS

Following the guidelines of Hwang et al. (2020) and Lee et al. (2021), the prediction outcomes of the classifiers were validated by employing a confusion matrix (Table 1): precision [equation (5)], recall [equation (6)], accuracy [equation (7)] and F1-score [equation (8)]. As shown in outcomes from "before COVID-19" dataset, the Logistic regression classifier model using the TF-IDF achieved the highest accuracy with 89.07%. Additionally, XGBoost with BoW (88.83%) achieved relatively high accuracy. As reported in results from "after COVID-19" dataset, XGBoost with BoW achieved

Figure 2. Flow chart of the study method

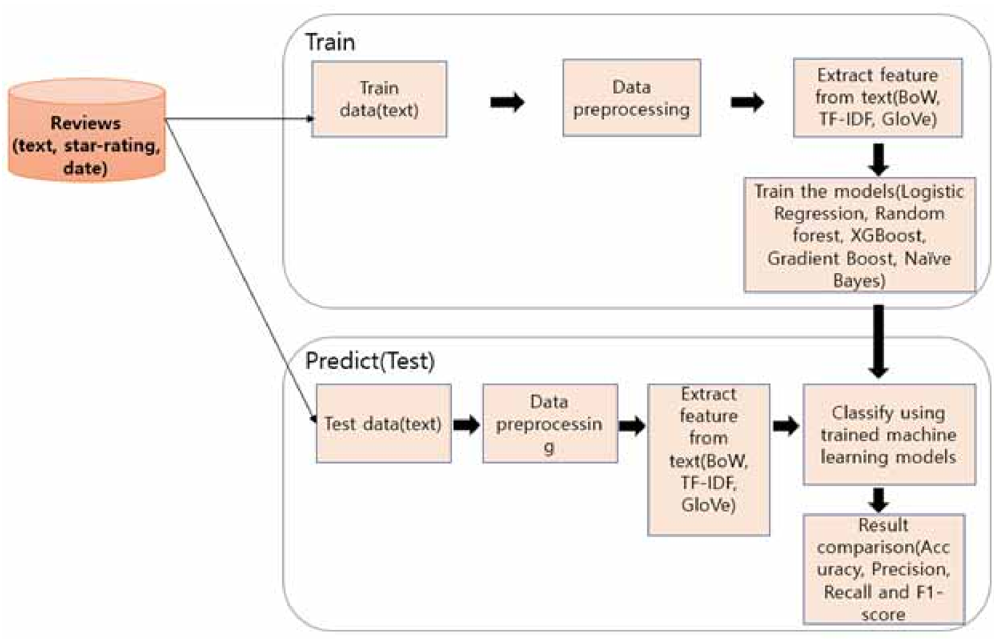


Table 1. Confusion matrix

Confusion matrix		True class	
		Positive	**Negative**
Predictive class	**Positive**	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

the highest accuracy with 89.16%. Moreover, Logistic regression with BoW (89.10%) achieved the relatively high accuracy. The summary of the accuracy results is reported in Figure 3, 4 and Table 2, 3.

$$Precision = \frac{TP}{\left(TP + FP\right)} \tag{5}$$

$$Recall = \frac{TP}{\left(TP + FN\right)} \tag{6}$$

$$Accuracy = \frac{\left(TP + TN\right)}{\left(TP + TN + FP + FN\right)} \tag{7}$$

Table 2. Before Covid-19 Results of each classifier

Model	Embedding	Class	Accuracy	Precision	Recall	F1-score
Logistic Regression	BoW	pos	0.8831	0.8506	0.9295	0.88828
		neg		0.9223	0.8367	0.87741
	TF-IDF	pos	0.8908	0.877	0.9091	0.89273
		neg		0.9056	0.8725	0.88873
	GloVe	pos	0.8658	0.8559	0.8796	0.86762
		neg		0.8762	0.8519	0.86389
XGBoost	BoW (‘learning_rate’: 0.1, ‘max_depth’: 8, ‘n_estimators’: 1000)	pos	0.8884	0.8606	0.9269	0.89252
		neg		0.9208	0.8498	0.8839
	TF-IDF (‘learning_rate’: 0.1, ‘max_depth’: 6, ‘n_estimators’: 1000)	pos	0.8879	0.8745	0.9057	0.88984
		neg		0.9022	0.8701	0.88585
	GloVe (‘learning_rate’: 0.05, ‘max_depth’: 4, ‘n_estimators’: 1000)	pos	0.7713	0.7886	0.7414	0.76428
		neg		0.756	0.8012	0.77797
Random Forest	BoW (‘max_depth’: 50, ‘min_samples_leaf’: 1, ‘min_samples_split’: 2, ‘n_estimators’: 1000)	pos	0.8618	0.8314	0.9077	0.86783
		neg		0.8983	0.8159	0.85512
	TF-IDF (‘max_depth’: 50, ‘min_samples_leaf’: 1, ‘min_samples_split’: 2, ‘n_estimators’: 600)	pos	0.8676	0.8697	0.8647	0.86723
		neg		0.8655	0.8705	0.86799
	GloVe (‘max_depth’: 50, ‘min_samples_leaf’: 2, ‘min_samples_split’: 2, ‘n_estimators’: 600)	pos	0.7386	0.8299	0.6002	0.69662
		neg		0.6869	0.877	0.77037
GBM	BoW (‘learning_rate’: 0.1, ‘max_depth’: 8, ‘n_estimators’: 1000)	pos	0.8868	0.8591	0.9253	0.89095
		neg		0.919	0.8483	0.88222
	TF-IDF (‘learning_rate’: 0.2, ‘max_depth’: 8, ‘n_estimators’: 1000)	pos	0.8858	0.8743	0.901	0.88749
		neg		0.8979	0.8705	0.884
	GloVe (‘learning_rate’: 0.05, ‘max_depth’: 6, ‘n_estimators’: 200)	pos	0.6914	0.6466	0.8446	0.73242
		neg		0.776	0.5383	0.63564
Naïve Bayes	BoW	pos	0.8001	0.7488	0.9033	0.81881
		neg		0.8781	0.6969	0.77712
	TF-IDF	pos	0.7413	0.7039	0.8331	0.76307
		neg		0.7956	0.6496	0.71523
	GloVe	pos	0.8037	0.8066	0.799	0.80277
		neg		0.8009	0.8084	0.80462

Table 3. After Covid-19 Results of each classifier

Model	Embedding	Class	Accuracy	Precision	Recall	F1-score
Logistic Regression	BoW	pos	0.891	0.9002	0.8796	0.88978
		neg		0.8823	0.9025	0.89227
	TF-IDF	pos	0.8867	0.9257	0.841	0.88131
		neg		0.8543	0.9325	0.89169
	GloVe	pos	0.8575	0.9099	0.7936	0.84778
		neg		0.817	0.9214	0.86607
XGBoost	BoW (‘learning_rate’: 0.2, ‘max_depth’: 6, ‘n_estimators’: 1000)	pos	0.8916	0.9146	0.8639	0.88856
		neg		0.8711	0.9194	0.89457
	TF-IDF (‘learning_rate’: 0.1, ‘max_depth’: 4, ‘n_estimators’: 1000)	pos	0.8842	0.9244	0.8369	0.87849
		neg		0.851	0.9316	0.88948
	GloVe (‘learning_rate’: 0.1, ‘max_depth’: 6, ‘n_estimators’: 200)	pos	0.7656	0.8335	0.664	0.73912
		neg		0.7208	0.8673	0.78728
Random Forest	BoW (‘max_depth’: 50, ‘min_samples_leaf’: 1, ‘min_samples_split’: 2, ‘n_estimators’: 600)	pos	0.8662	0.8997	0.8243	0.86033
		neg		0.8379	0.9081	0.87156
	TF-IDF (‘max_depth’: 50, ‘min_samples_leaf’: 1, ‘min_samples_split’: 10, ‘n_estimators’: 1000)	pos	0.8581	0.927	0.7773	0.84559
		neg		0.8083	0.9388	0.86865
	GloVe (‘max_depth’: 20, ‘min_samples_leaf’: 2, ‘min_samples_split’: 2, ‘n_estimators’: 1000)	pos	0.6988	0.9088	0.442	0.59476
		neg		0.6314	0.9556	0.76036
GBM	BoW (‘learning_rate’: 0.2, ‘max_depth’: 6, ‘n_estimators’: 1000)	pos	0.8896	0.908	0.867	0.88703
		neg		0.8727	0.9122	0.89203
	TF-IDF (‘learning_rate’: 0.1, ‘max_depth’: 6, ‘n_estimators’: 1000)	pos	0.8815	0.9232	0.8322	0.87531
		neg		0.8472	0.9308	0.88703
	GloVe (‘learning_rate’: 0.05, ‘max_depth’: 4, ‘n_estimators’: 200)	pos	0.6971	0.6619	0.8057	0.72676
		neg		0.7518	0.5884	0.66014
Naïve Bayes	BoW	pos	0.8234	0.8079	0.8485	0.82771
		neg		0.8405	0.7982	0.81881
	TF-IDF	pos	0.7735	0.7664	0.786806	0.77645
		neg		0.781	0.7601	0.7704
	GloVe	pos	0.7662	0.9063	0.5937	0.71747
		neg		0.6979	0.9386	0.80058

Figure 3. Summary of prediction accuracy before Covid-19

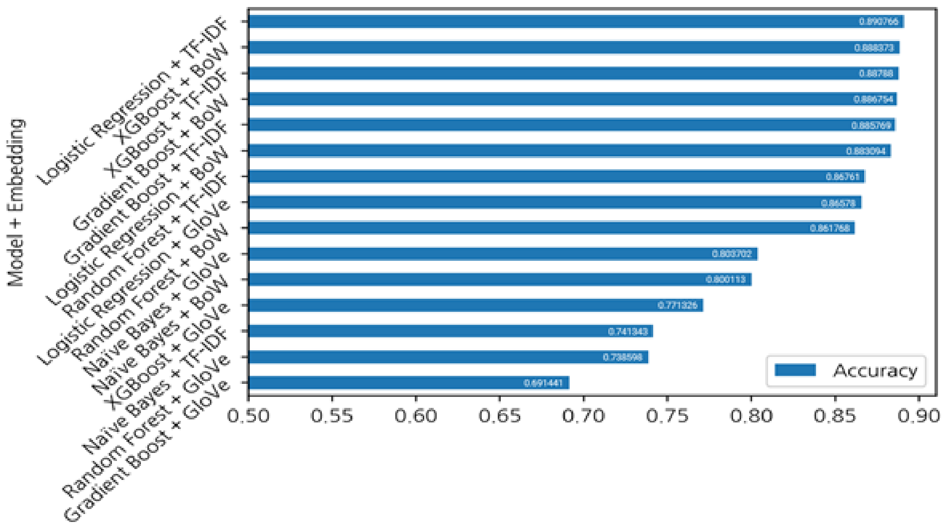
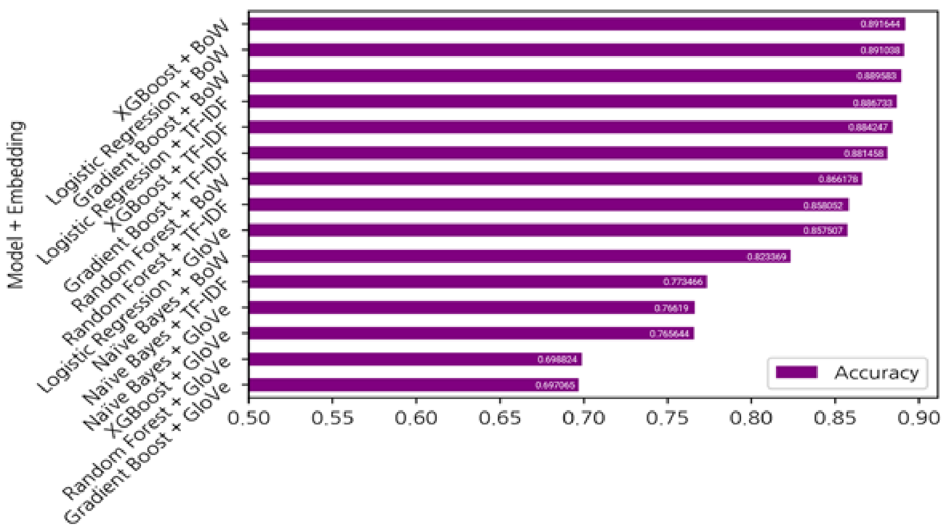


Figure 4. Summary of prediction accuracy after Covid-19



$$F1 - score = \frac{2 \times \left(Recall \times Precision\right)}{\left(Recall + Precision\right)} \tag{8}$$

DISCUSSION

Rare studies predicted the satisfaction of healthcare service users through comparing cutting-edge machine learning algorithms. The present study seeks to find a machine learning model that predicts user satisfaction with healthcare services using review data and satisfaction ratings from Google play store.

Along with the swift diffusion of COVID-19, the mobile healthcare service market is thriving again (Coherent Market Insights, 2021). With this trend, several companies have been making efforts to improve their mobile healthcare services for better user experiences (Big Innovation Centre, 2021). Furthermore, considering that user satisfaction is a significant element in the success of a marketing strategy, several studies have tried to explore user satisfaction in the context of mobile healthcare services. Thus, large scale online reviews were computed in this study using several machine learning approaches (i.e., logistic regression, random forest, Naïve Bayes, XGBoost, and GBM) with feature representation techniques (i.e., TF-IDF, BoW, GloVe) for predicting user satisfaction of healthcare services. Based on the outcomes from "before COVID-19" dataset, Logistic regression with TF-IDF achieved the highest accuracy of 89.07%. Moreover, as indicated in the results from "after COVID-19" dataset, XGBoost with BoW achieved the highest accuracy of 89.16%. The prediction accuracies achieved in this study can be considered superior to the results of existing studies (Hwang et al., 2020; La et al., 2021). Overall, Logistic regression with TF-IDF and XGBoost with BoW are proposed to have better performances in predicting user satisfaction for healthcare services.

For practical application, this research has the following implications. As shown in the results, most classifiers with GloVe have low accuracy. This outcome is different from the viewpoints of Lim et al. (2021) and Krouska et al. (2020), which implies that GloVe is suitable for extracting text features when using machine learning methods. This means that GloVe may not be appropriate for analyzing online reviews of mobile healthcare service. Thus, healthcare service providers need not consider applying GloVe, instead of TF-IDF or BoW, for predicting user satisfaction.

Furthermore, based on the outcomes from "after COVID-19" dataset, this research found that XGBoost using BoW shows similar predictive performance to Logistic regression using BoW. Thus, the following suggestions are proposed for healthcare service providers in the context of COVID-19 pandemic. The use of XGBoost with BoW can be prioritized for developing online review analysis platforms for healthcare service providers; it showed better predictive performance in predicting user satisfaction. However, an advanced model such as XGBoost need not be introduced when a company lacks human or financial resources (see Nan et al., 2022). This is because, Logistic regression using BoW also exhibits sufficient prediction performance."

CONCLUSION

Existing studies (e.g., Zhang et al., 2017; Keikhosrokiani et al., 2020) on user experiences of healthcare services used a limited number of samples and traditional approaches (e.g., SEM). For addressing the limitations of previous studies, the current research computed large scale online reviews using machine learning approaches for exploring user satisfaction with healthcare services. As a result, in this research, machine learning techniques showed high accuracy of roughly 90% in predicting user satisfaction, implying that machine learning approaches are effective methods for exploring user experience and satisfaction. In addition, considering that online reviews are easy to obtain and machine learning techniques can effectively process such a large amount of data, the approach used in this research saves money and time (see Kim et al., 2021) compared to survey methods that have been widely used to explore user experience (Ju and Zhang, 2020; Keikhosrokiani et al., 2020; Nan et al., 2022). Overall, the current research can contribute to extending the literature of user experience of healthcare services in the dimensions of big data and machine learning approaches.

In terms of theoretical implications, the results show that user-generated texts such as online reviews can be utilized to predict user satisfaction. This viewpoint is supported by Kim et al. (2021) and Cheng and Jin (2019), which implies that online reviews and user experiences have a close association. This is because online reviews generally contain user's feelings or perceptions after using specific services or products (Kim et al., 2021; Park et al. 2021; Zhao et al., 2019). Furthermore, this finding can be explained by affect theory (Kratzwald et al., 2018). The theory implies that when users are satisfied with a particular service, they will write online reviews with positive emotions and praise the service provider, and conversely, dissatisfaction will lead them to write online reviews with negative emotions and criticize the service provider (Xu, 2020). Therefore, this study also can contribute to expanding the literature of affect theory in the aspect of computational approach.

The current research has the following limitations that should be addressed in the future. First, this research only employed one mobile healthcare application (i.e., Samsung Health) for detecting user satisfaction. Therefore, it is necessary to validate the methodology in the context of other mobile healthcare applications. Second, Google play store does not provide users' demographic information; therefore, the research did not consider age or gender in our approaches. Third, this research only analyzed English data; thus, future studies should compute other languages for validating our approaches. Fourth, although deep learning approach also is widely adopted in text classification area, this study did not consider the approach because deep learning is out of the scope of the present research. However, future research can employ some deep learning methods such as Bidirectional Encoder Representations from Transformers (BERT) to predicting user satisfaction in the context of mobile healthcare services.

ACKNOWLEDGMENT

This work was supported by a National Research Foundation of Korea (NRF) grant funded by the Korean government (NRF-2020R1A2C1014957).

REFERENCES

Akter, S., D'Ambra, J., & Ray, P. (2013a). Development and validation of an instrument to measure user perceived service quality of mHealth. *Information & Management*, *50*(4), 181–195. doi:10.1016/j.im.2013.03.001

Akter, S., Ray, P., & D'Ambra, J. (2013b). Continuance of mHealth services at the bottom of the pyramid: The roles of service quality and trust. *Electronic Markets*, *23*(1), 29–47. doi:10.1007/s12525-012-0091-5

Alam, M. M. D., Alam, M. Z., Rahman, S. A., & Taghizadeh, S. K. (2021). Factors influencing mHealth adoption and its impact on mental well-being during COVID-19 pandemic: A SEM-ANN approach. *Journal of Biomedical Informatics*, *116*, 103722. doi:10.1016/j.jbi.2021.103722 PMID:33705856

Alanzi, T. M., Althumairi, A., Aljaffary, A., Alfayez, A., Alsalman, D., Alanezi, F., & AlThani, B. et al. (2021). Evaluation of the Mawid mobile healthcare application in delivering services during the COVID-19 pandemic in Saudi Arabia. *International Health*. PMID:33864074

Bansal, B., & Srivastava, S. (2018). Sentiment classification of online consumer reviews using word vector representations. *Procedia Computer Science*, *132*, 1147–1153. doi:10.1016/j.procs.2018.05.029

Bhattacherjee, A. (2001). Understanding information systems continuance: An expectation-confirmation model. *Management Information Systems Quarterly*, *25*(3), 351–370. doi:10.2307/3250921

Big Innovation Centre. (2021). *How mobile phones and wearable technology are transforming healthcare - New Statesman*. Available on: https://www.newstatesman.com/spotlight/healthcare/2021/09/how-mobile-phones-and-wearable-technology-are-transforming-healthcare

Bol, N., Helberger, N., & Weert, J. C. (2018). Differences in mobile health app use: A source of new digital inequalities? *The Information Society*, *34*(3), 183–193. doi:10.1080/01972243.2018.1438550

Budholiya, K., Shrivastava, S. K., & Sharma, V. (2020). An optimized XGBoost based diagnostic system for effective prediction of heart disease. *Journal of King Saud University-Computer and Information Sciences*.

Chatterjee, S., Goyal, D., Prakash, A., & Sharma, J. (2021). Exploring healthcare/health-product ecommerce satisfaction: A text mining and machine learning application. *Journal of Business Research*, *131*, 815–825. doi:10.1016/j.jbusres.2020.10.043

Chen, C., Liaw, A., & Breiman, L. (2004). Using random forest to learn imbalanced data. University of California, Berkeley.

Cheng, M., & Jin, X. (2019). What do Airbnb users care about? An analysis of online review comments. *International Journal of Hospitality Management*, *76*, 58–70. doi:10.1016/j.ijhm.2018.04.004

Cheng, Y., & Jiang, H. (2020). How do AI-driven chatbots impact user experience? Examining gratifications, perceived privacy risk, satisfaction, loyalty, and continued use. *Journal of Broadcasting & Electronic Media*, *64*(4), 592–614. doi:10.1080/08838151.2020.1834296

Coherent Market Insights. (2021). *Digital Healthcare Market to grow US$ 7,304.3 Mn by end of 2028, Says Coherent Market Insights*. Available in: https://www.prnewswire.com/news-releases/digital-healthcare-market-to-grow-us-7-304-3-mn-by-end-of-2028--says-coherent-market-insights-301434208.html

Global Market Insights. (2019). *At 38% CAGR, mHealth Market to Hit $289.4 Billion by 2025*. Global Market Insights, Inc. Available in: https://www.prnewswire.com/news-releases/at-38-cagr-mhealth-market-to-hit-289-4-billion-by-2025-global-market-insights-inc-300802727.html

Hakak, S., Alazab, M., Khan, S., Gadekallu, T. R., Maddikunta, P. K. R., & Khan, W. Z. (2021). An ensemble machine learning approach through effective feature extraction to classify fake news. *Future Generation Computer Systems*, *117*, 47–58. doi:10.1016/j.future.2020.11.022

Handayani, P. W., Meigasari, D. A., Pinem, A. A., Hidayanto, A. N., & Ayuningtyas, D. (2018). Critical success factors for mobile health implementation in Indonesia. *Heliyon*, *4*(11), e00981. doi:10.1016/j.heliyon.2018.e00981 PMID:30519665

Hew, K. F., Hu, X., Qiao, C., & Tang, Y. (2020). What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach. *Computers & Education*, *145*, 103724. doi:10.1016/j.compedu.2019.103724

Ho-Dac, N. N., Carson, S. J., & Moore, W. L. (2013). The effects of positive and negative online customer reviews: Do brand strength and category maturity matter? *Journal of Marketing*, *77*(6), 37–53. doi:10.1509/jm.11.0011

Hu, N., Zhang, T., Gao, B., & Bose, I. (2019). What do hotel customers complain about? Text analysis using structural topic model. *Tourism Management*, *72*, 417–426. doi:10.1016/j.tourman.2019.01.002

Hwang, S., Kim, J., Park, E., & Kwon, S. J. (2020). Who will be your next customer: A machine learning approach to customer return visits in airline services. *Journal of Business Research*, *121*, 121–126. doi:10.1016/j.jbusres.2020.08.025

Imaginovation. (2021). *Developing a Healthcare App in 2021: What Do Patients Really Want?* Available in: https://imaginovation.net/blog/developing-a-mobile-health-app-what-patients-really-want/

Jaana, M., & Paré, G. (2020). Comparison of mobile health technology use for Self-Tracking between older adults and the general adult population in Canada: Cross-sectional survey. *JMIR mHealth and uHealth*, *8*(11), e24718. doi:10.2196/24718 PMID:33104517

Jaiswal, K., & Anand, V. (2021). A Survey on IoT-Based Healthcare System: Potential Applications, Issues, and Challenges. In *Advances in Biomedical Engineering and Technology* (pp. 459–471). Springer. doi:10.1007/978-981-15-6329-4_38

Jang, J., & Yi, M. Y. (2017, May). Modeling user satisfaction from the extraction of user experience elements in online product reviews. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1718-1725). doi:10.1145/3027063.3053097

Jing, L. P., Huang, H. K., & Shi, H. B. (2002, November). Improved feature selection approach TFIDF in text mining. *Proceedings - International Conference on Machine Learning and Cybernetics*, *2*, 944–946. doi:10.1109/ICMLC.2002.1174522

Joachims, T. (1997). A probabilistic analysis of the rocchio algorithm with TF-IDF for text categorization. In *Proceedings of the 14th International Conference on Machine Learning*. Morgan Kaufmann Publishers Inc.

Jones, N. R., Qureshi, Z. U., Temple, R. J., Larwood, J. P., Greenhalgh, T., & Bourouiba, L. (2020). Two metres or one: what is the evidence for physical distancing in covid-19? *BMJ, 370*.

Ju, C., & Zhang, S. (2020). Research on user' continuous usage of online healthcare services from the perspective of affect appeal. *Journal of Technology in Behavioral Science*, *5*(3), 215–225. doi:10.1007/s41347-020-00128-9

Karimi, S. S., Saintilan, N., Wen, L., & Valavi, R. (2019). Application of machine learning to model wetland inundation patterns across a large semiarid floodplain. *Water Resources Research*, *55*(11), 8765–8778. doi:10.1029/2019WR024884

Keikhosrokiani, P., Mustaffa, N., Zakaria, N., & Abdullah, R. (2020). Assessment of a medical information system: The mediating role of use and user satisfaction on the success of human interaction with the mobile healthcare system (iHeart). *Cognition Technology and Work*, *22*(2), 281–305. doi:10.1007/s10111-019-00565-4

Khalil, A. A., Hidayanto, A. N., & Prabowo, H. (2020, November). Identification of Factor Affecting Continuance Usage Intention of mHealth Application: A Systematic Literature Review. In *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)* (pp. 1-6). IEEE. doi:10.1109/ICICoS51170.2020.9299038

Kim, J. H., Nan, D., Kim, Y., & Park, M. H. (2021). Computing the User Experience via Big Data Analysis: A Case of Uber Services. *CMC-Computers Materials & Continua*, *67*(3), 2819–2829. doi:10.32604/cmc.2021.014922

Konstantinov, A. V., & Utkin, L. V. (2021). Interpretable machine learning with an ensemble of gradient boosting machines. *Knowledge-Based Systems*, *222*, 106993. doi:10.1016/j.knosys.2021.106993

Kratzwald, B., Ilić, S., Kraus, M., Feuerriegel, S., & Prendinger, H. (2018). Deep learning for affective computing: Text-based emotion recognition in decision support. *Decision Support Systems*, *115*, 24–35. doi:10.1016/j.dss.2018.09.002

Krouska, A., Troussas, C., & Virvou, M. (2020). Deep learning for twitter sentiment analysis: the effect of pre-trained word embedding. In *Machine Learning Paradigms* (pp. 111–124). Springer. doi:10.1007/978-3-030-49724-8_5

Kumar, V., Dalla Pozza, I., & Ganesh, J. (2013). Revisiting the satisfaction–loyalty relationship: Empirical generalizations and directions for future research. *Journal of Retailing*, *89*(3), 246–262. doi:10.1016/j.jretai.2013.02.001

La, L., Xu, F., Ren, Q., Zhen, F., & Lobsang, T. (2021). Understanding Consumers' Sentiment Expressions in Online Reviews: A Hybrid Approach. *Journal of International Consumer Marketing*, 1–15. doi:10.1080/08961530.2021.1982807

Lee, S., Ji, H., Kim, J., & Park, E. (2021). What books will be your bestseller? A machine learning approach with Amazon Kindle. *The Electronic Library*, *39*(1), 137–151. doi:10.1108/EL-08-2020-0234

Lee, W. I., & Figueredo, N. M. (2021). Exploring the Perspective of Service Quality in mHealth Services during the COVID-19 Pandemic. *International Journal of Economics and Management Engineering*, *15*(5), 562–566.

Li, C., & Zhang, J. (2014). *Prediction of yelp review star rating using sentiment analysis*. Stanford CEE.

Lim, S., Prade, H., & Richard, G. (2021). Classifying and completing word analogies by machine learning. *International Journal of Approximate Reasoning*, *132*, 1–25. doi:10.1016/j.ijar.2021.02.002

Lin, W., Xu, M., He, J., & Zhang, W. (2021). Privacy, security and resilience in mobile healthcare applications. *Enterprise Information Systems*, 1–15. doi:10.1080/17517575.2021.1939896

Liu, Z., Qi, M., Shen, C., Fang, Y., & Zhao, X. (2021). Cascade saccade machine learning network with hierarchical classes for traffic sign detection. *Sustainable Cities and Society*, *67*, 102700. doi:10.1016/j.scs.2020.102700

Luo, Y., & Xu, X. (2021). Comparative study of deep learning models for analyzing online restaurant reviews in the era of the COVID-19 pandemic. *International Journal of Hospitality Management*, *94*, 102849. doi:10.1016/j.ijhm.2020.102849 PMID:34785843

Meng, F., Guo, X., Zhang, X., Peng, Z., & Lai, K. H. (2020, January). Examining the role of technology anxiety and health anxiety on elderly users' continuance intention for mobile health services use. *Proceedings of the 53rd Hawaii International Conference on System Sciences*. doi:10.24251/HICSS.2020.403

Ming, L. C., Untong, N., Aliudin, N. A., Osili, N., Kifli, N., Tan, C. S., Goh, K. W., Ng, P. W., Al-Worafi, Y. M., Lee, K. S., & Goh, H. P. (2020). Mobile health apps on COVID-19 launched in the early days of the pandemic: Content analysis and review. *JMIR mHealth and uHealth*, *8*(9), e19796. doi:10.2196/19796 PMID:32609622

Nan, D., Kim, Y., Park, M. H., & Kim, J. H. (2020). What motivates users to keep using social mobile payments? *Sustainability*, *12*(17), 6878. doi:10.3390/su12176878

Nan, D., Lee, H., Kim, Y., & Kim, J. H. (2022). My video game console is so cool! A coolness theory-based model for intention to use video game consoles. *Technological Forecasting and Social Change*, *176*, 121451. doi:10.1016/j.techfore.2021.121451 PMID:34980931

Nandanwar, A. K., & Choudhary, J. (2021). Semantic Features with Contextual Knowledge-Based Web Page Categorization Using the GloVe Model and Stacked BiLSTM. *Symmetry*, *13*(10), 1772. doi:10.3390/sym13101772

Oppong, E., Hinson, R. E., Adeola, O., Muritala, O., & Kosiba, J. P. (2021). The effect of mobile health service quality on user satisfaction and continual usage. *Total Quality Management & Business Excellence*, *32*(1-2), 177–198. doi:10.1080/14783363.2018.1541734

Park, G., Nan, D., Park, E., Kim, K. J., Han, J., & del Pobil, A. P. (2021, January). Computers as Social Actors? Examining How Users Perceive and Interact with Virtual Influencers on Social Media. In *2021 15th International Conference on Ubiquitous Information Management and Communication (IMCOM)* (pp. 1-6). IEEE. doi:10.1109/IMCOM51814.2021.9377397

Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543). doi:10.3115/v1/D14-1162

Saheb, T. (2020). An empirical investigation of the adoption of mobile health applications: Integrating big data and social media services. *Health and Technology*, *10*(5), 1063–1077. doi:10.1007/s12553-020-00422-9

Sakketou, F., & Ampazis, N. (2020). A constrained optimization algorithm for learning GloVe embeddings with semantic lexicons. *Knowledge-Based Systems*, *195*, 105628. doi:10.1016/j.knosys.2020.105628

Samsung Newsroom. (2020). Available in: https://news.samsung.com/ kr/%EC%8 2%BC%EC %84%B1-%EC %8A%A4%E B%A7%88% ED%8A%B 8-tv%EC% 84%9C-%EC%82%BC%EC%84%B1-%ED%97%AC%EC%8A%A4-%EC% 84%9C%EB %B9%84 %EC%8A %A4-%EA% B0%9C%E C%8B%9C

Schaeffer, S. E., & Sanchez, S. V. R. (2020). Forecasting client retention—A machine-learning approach. *Journal of Retailing and Consumer Services*, *52*, 101918. doi:10.1016/j.jretconser.2019.101918

Shah, B. K., Jaiswal, A. K., Shroff, A., Dixit, A. K., Kushwaha, O. N., & Shah, N. K. (2021, January). Sentiments Detection for Amazon Product Review. In *2021 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-6). IEEE.

Soumya, S., & Pramod, K. V. (2020). Sentiment analysis of malayalam tweets using machine learning techniques. *ICT Express*, *6*(4), 300–305. doi:10.1016/j.icte.2020.04.003

Wang, X., & Goh, D. H. L. (2020). Components of game experience: An automatic text analysis of online reviews. *Entertainment Computing*, *33*, 100338. doi:10.1016/j.entcom.2019.100338

Xu, X. (2020). Examining consumer emotion and behavior in online reviews of hotels when expecting managerial response. *International Journal of Hospitality Management*, *89*, 102559. doi:10.1016/j.ijhm.2020.102559

Xu, Y., Wu, G., & Chen, Y. (2022). Predicting Patients' Satisfaction With Doctors in Online Medical Communities: An Approach Based on XGBoost Algorithm. *Journal of Organizational and End User Computing*, *34*(4), 1–17. doi:10.4018/JOEUC.287571

Yang, H. E., Wu, C. C., & Wang, K. C. (2009). An empirical analysis of online game service satisfaction and loyalty. *Expert Systems with Applications*, *36*(2), 1816–1825. doi:10.1016/j.eswa.2007.12.005

Yonhap News. (2021). *More than 200 million people used Samsung's fitness app in 2020*. Available in: https://en.yna.co.kr/view/AEN20210106002100320

Zahoor, K., Bawany, N. Z., & Hamid, S. (2020, November). Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning. In *2020 21st International Arab Conference on Information Technology (ACIT)* (pp. 1-6). IEEE. doi:10.1109/ACIT50332.2020.9300098

Zhang, X., Han, X., Dang, Y., Meng, F., Guo, X., & Lin, J. (2017). User acceptance of mobile health services from users' perspectives: The role of self-efficacy and response-efficacy in technology acceptance. *Informatics for Health & Social Care*, *42*(2), 194–206. doi:10.1080/17538157.2016.1200053 PMID:27564428

Zhang, X., Yan, X., Cao, X., Sun, Y., Chen, H., & She, J. (2018). The role of perceived e-health literacy in users' continuance intention to use mobile healthcare applications: An exploratory empirical study in China. *Information Technology for Development*, *24*(2), 198–223. doi:10.1080/02681102.2017.1283286

Zhao, Y., Ni, Q., & Zhou, R. (2018). What factors influence the mobile health service adoption? A meta-analysis and the moderating role of age. *International Journal of Information Management*, *43*, 342–350. doi:10.1016/j.ijinfomgt.2017.08.006

Zhao, Y., Xu, X., & Wang, M. (2019). Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews. *International Journal of Hospitality Management*, *76*, 111–121. doi:10.1016/j.ijhm.2018.03.017

Zheng, Y., Kong, H., Petzhold, G., Barcelos, M. M., Zegras, C. P., & Zhao, J. (2021). *User Satisfaction and Service Quality Improvement Priority of Bus Rapid Transit in Belo Horizonte*. Case Studies on Transport Policy. doi:10.1016/j.cstp.2021.10.011

Haein Lee is MS/PhD student at Department of Applied Artificial Intelligence/Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University. Her research focuses on Natural Language Processing, Deep Learning, and Machine Learning.

Seon Hong Lee is MS/PhD student at Department of Applied Artificial Intelligence/Department of Human-Artificial Intelligence Interaction, Sungkyunkwan University. His research focuses on Natural Language Processing, Deep Learning, and Machine Learning.

Dongyan Nan is a PhD student at Department of Human-Artificial Intelligence Interaction/Department of Interaction Science, Sungkyunkwan University. His research focuses on user experience, big data, and technology management. He has authored or co-authored 10 papers in major conferences and journals. Dongyan Nan is the corresponding author and can be contacted via ndyzxy0926@skku.edu.

Jang Hyun Kim is an associate professor at Department of Human-Artificial Intelligence Interaction/Department of Interaction Science/Department of Applied Artificial Intelligence, Sungkyunkwan University. His research focuses on social/semantic data analysis, social media, and future media. He has authored over 50 papers in major journals such as Information Processing & Management, Telematics & Informatics, Cities, Government Information Quarterly, Technological Forecasting and Social Change, and Journal of Computer-Mediated Communication. Jang Hyun Kim is the corresponding author and can be contacted via alohakim@skku.edu.