Modelling Customer Satisfaction from online reviews using Interpretative Machine Learning and Kano model

1. Introduction

Customer satisfaction is a psychological state, derived from a customer's subjective evaluation of a product or service provided based on expectations and actual performance (Anderson, Fornell, and Lehmann 1994; Oliver 1980; Woodruff, Cadotte, and Jenkins 1983). Analyzing customer satisfaction is significantly critical to company's sales performance in that customer satisfaction facilitates product or service improvement and customer loyalty derives from customer satisfaction and it facilitates (Chris Baumann et al, 2012, Chiang, Guo, and Pai 2009; Hallowell 1996; Hui, Wan, and Ho 2007; Kim, Ng, and Kim 2009; Sepehr and Head 2018). For these reasons, modeling and measuring customer satisfaction in valid way is a very important research topic.

Generally, customer satisfaction is measured empirically through customer survey. However, survey methods require a lot of time and cost. And survey responses heavily rely on the format of questionnaire and the willingness of the respondents to participate. Moreover, the data obtained from surveys may quickly become outdated (Culotta and Culter 2016).

As an alternative way, the number of studies using online review to measure customer satisfaction is increased. In the era of Industry 4.0 and Web 2.0, customers' reviewing behavior is considered one of the customer purchasing processes (Chong et al. 2017; Grewal, Cote, and Baumgartner 2004; Liu, Bi, and Fan 2017). Online reviews represent VoC (voice of customer) since it contains not only customers' concerns, sentiments and opinions but also customer attributes (purchase date, region etc), which are valuable for product and marketing managers to understand customer response (Farhadloo, Patterson, and Rolland 2016; Guo, Barnes, and Jia 2017; Rpournarakis, Sotiropoulos, and Giaglis 2017; Qi et al. 2016). Furthermore, as the number of online reviews is huge enough to represent sample data's validity and reliability. Thus, utilizing online reviews on measuring customer satisfaction is worthwhile in terms of its attributes.

Reviewing pre-studies attempting to understand customer satisfaction from online reviews, Bi et al (2019) divided the studies into two categories, (1) mining the customer satisfaction dimensions (CSDs) from online reviews (Guo, Barnes, and Jia 2017; Tirunillai and Tellis 2014) and (2) modelling customer satisfaction from online reviews (Decker and Trusov 2010; Farhadloo, Patterson, and Rolland 2016; Qi et al. 2016; Xiao, Wei, and Dong 2016). The former mainly focuses on extracting the CSDs from online reviews and analyzing the relative importance of each CSD. On the other hand, the latter mainly focuses on analyzing the effects of customer sentiments toward product or service attributes on customer satisfaction.

These studies still have some limitations. Among studies mining CSDs from online reviews, there are no studies that attempted to measure the effects of CSDs on customer satisfaction are not measured except Bi et al (2019) (Guo, Barnes, and Jia 2017; Tirunillai and Tellis 2014). Besides, nearly all the existing studies assume that CSDs mined from online reviews satisfy additive independence, i.e. the online rating (customer satisfaction) assigned by a customer is a linear combination of the sentiment orientations of all the CSDs mentioned in the customer's review (Decker and Trusov 2010; Farhadloo, Pattersom, and Rollland 2016; Qi et al. 2016; Xiao, Wei, and Dong 2016). However, Bi et al (2019) considered that CSDs have complex relationships (e.g. multicollinearity and non-linearities) among different CSDs and the customer satisfaction since the CSDs are automatically mined from the online reviews whereas obtained from the designed questionnaires. In addition, in the existing studies on customer satisfaction, it has been verified that CSDs can be classified into different categories that would affect the customer satisfaction in different ways (Bi et al 2019; Chen and Chauang 2009; Ji et al. 2014; Li, Du, and Chin 2018; Kano et al. 1984; Matzler and Hinterhuber 1998).

The method and techniques used in mining and analyzing online reviews also have some limitations. The existing studies used unsupervised learning (SVM, k-means clustering) in classifying customer sentiment on product or service. However, generally predicting performance of unsupervised learning has lower performance than predicting performance of supervised learning. Also, machine learning models are more effective than deep learning based models in diagnosing and analyzing one-dimensional data (numerical data). Last not but not least, it is limited to only use big data techniques in fully and richly interpreting results. It is recommended to apply economical or managerial theories in interpreting results to extract managerial and academic implication.

To resolve the limitations of existing studies, the author established the goal of this study in the following objectives:

(1) To model the effects of customer sentiments (positive or negative) towards each CSD on customer satisfaction more reliable and valid way.
(2) To interpret the effect of CSDs in customer satisfaction applying economical and managerial theories.

2. A framework for modelling customer satisfaction from online reviews

In this section, a process for modeling and analyzing customer satisfaction from online reviews is demonstrated, as shown in Figure 1. Each step in the framework is illustrated as follows.

(1) Online review: An online review is a customer generated text that contains the opinions of the customer on a product or service.
(2) Customer Satisfaction: Customer satisfaction is a psychological state, derived from a customer's overall subjective evaluation of a product or service provided based on expectations and actual performance. Similar with the prior studies (Decker and Trusov 2010; Farhadloo, Patterson, and Rolland 2016; Guo, Barnes, and Jia 2017; Tirunillai and Tellis 2014), in this study we consider that the online rating evaluated by a customer represents the customer's satisfaction with a product or service.
(3) Customer Satisfaction dimension (CSD): Usually, customers evaluate their satisfaction with a product or service based on their perceptions of several important attributes. The vocabularies that describe an important attribute of a product or service are defined as a CSD (Guo, Barnes, and Jia 2017; Tirunillai and Tellis 2014).
(4) Shapley value: The Shapley value is a solution concept used in game theory which is the average marginal contribution of a feature value across all possible coalitions.
(5) The category of the CSD: In the existing studies on customer satisfaction, it has been verified that CSDs can

be classified into different categories. In this study, according to the Kano model (Kano et al. 1984), the CSDs are classified into five categories (i.e. performance CSD, excitement CSD, must-be CSD, reverse CSD and indifferent CSD), as shown in Figure 2.

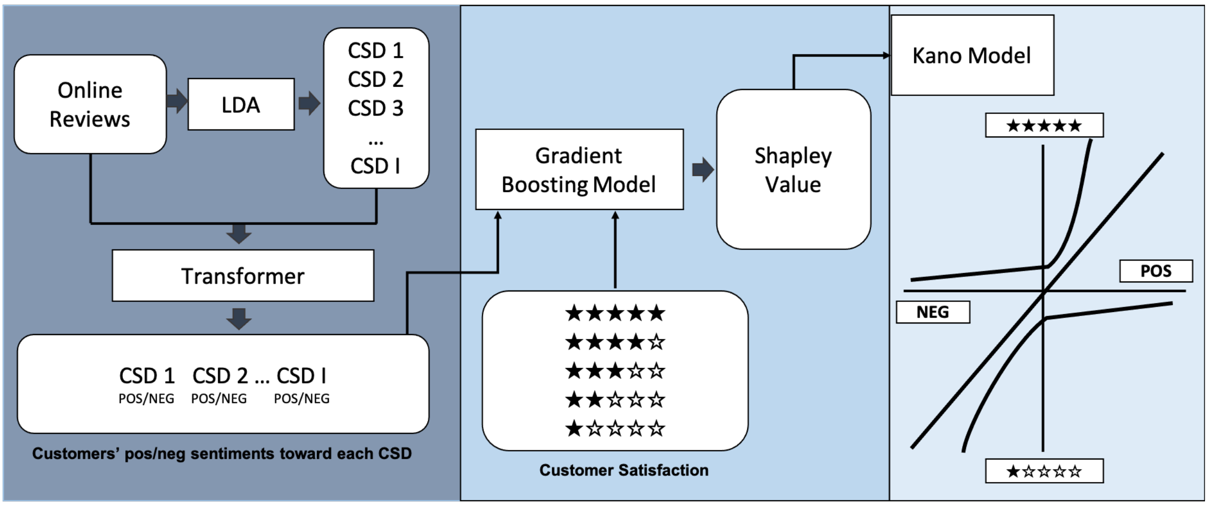


Figure 1. A framework for modelling customer satisfaction from online reviews.

Figure 2. Classification of CSDs in the Kano model.

Based on figure 1 and the descriptions of the basic concepts, the framework is mainly divided into three steps.

Part 1: Extracting topics and sentiments toward topics per one online review.

Part 2: Measuring the effects of customer sentiments toward each CSD on customer satisfaction.

Part 3: Analyzing each CSD's effects based on Shapley value and Kano model.

3. Method

3.1 Mining customers' sentiments toward CSDs from online reviews

As online reviews are in the form of written words, they need natural language processing (NLP) to use them for analysis (Cambria et al. 2017). In NLP, the unstructured data are transformed into the structured form that can be directly used for modelling customer satisfaction. In every step, different NLP methods are applied.

3.1.1. Extracting CSDs from online reviews based on LDA

To identify CSDs from online reviews, LDA was used to extract topics which are the product or service features that customers mainly consider. LDA is an unsupervised generative probabilistic model that can be used to effectively extract the latent topics from a large number of online reviews (Bi et al, 2019; Guo, Barnes, and Jia 2017; Blei, Ng, and Jordan 2003; Tirunillai and Tellis 2014). The topic here is a series of related words that describe one concept or one aspect. In the LDA model, each online review is thought of a s a mixed probability distribution composed of some topics, where each topic has a probability distribution on some words. The process of generating an online review is to repeatedly select a topic based on the topic distribution, and then select the words based on the probability distribution on words of each selected topic (Blei, Ng, and Jordan 2003). According to the process, once an online review has been generated, the probability of each word in the review can be expressed as a probability formula:

$$p(word|review) = \sum_{topic} p(word|topic) \times p(topic|review)$$

There were several data preprocessing steps to extract product related topic. First, all reviews were split into vocabularies. Second, emotional expressions, stopwords, negation words, brand name, and product name were excluded which don't take into account extracting CSDs. Third, lemmatization which makes vocabularies returned to the base or dictionary form of a word. Fourth, only reviews with one vocabulary were excluded. Fifth, authors composed a dictionary that concludes every vocabulary in whole reviews. Sixth, each vocabulary in the dictionary was numbered.

Let $R_i = \{r_{i1}^l, \ldots\ldots, r_{ic}^k, \ldots\ldots, r_{iC_i}^g\}$ denote the set of reviews with respect to the ith CSD in the set of online revies R, where r_{ic}^k denotes the cth review in R_i and the kth review in R, C_i is the total number of reviews concerning the ith CSD, l, k, g ϵ {1, 2, …, M}. To obtain R_i, the online reviews in set R are firstly divided into sentences based on the punctuations. By dividing into sentences, more than one topic can be described in one review. Then, according to the obtained topic clusters from LDA, R_i can be obtained by extracting the sentences that contain one of words in topic clusters.

3.1.2. Identifying the sentiment orientations of the reviews regarding each CSD using Transformer

In this subsection, every CSDs in each review are classified into positive or negative sentiment by using Transformer model. A transformer is a deep learning model that adopts the mechanism of attention, weighing the influence of different parts of the input data. **Advan**

As Transformer is one of supervised learning models, review data are separated into train set and test set. It needs train set which contains answer (sentiments). **Positional Encoding** Decision-makers constructed train set by identify sentiments of reviews in train set manually. After train set is prepared, decision-makers learn Transformer model with k-fold cross-validation and parameter tuning. And sentiments in test sets are identified based on established Transformer model.

According to the sentiment orientation of each review r_m, where * ϵ {Pos, Neg}, i = 1,2, …, I, m = 1,2,….,M. The results can be converted to nominally coded data, as shown in Table 1. In Table 1, 'Missing value' indicates no response concerning the CSD.

Let * denote the sentiment orientation of f_i in online review r_m, where * ϵ {Pos, Neg}, i = 1,2, …, I, m = 1,2,….,M. The results can be converted into structured data, as shown in Table 1. It can be seen from Table 1, if the sentiment orientation of f_i in online review r_m is positive, then $S_{im} = 1$; if the sentiment orientation is negative, then $S_i \ \ = -1$; if there is no review on f_i in online review r_m, then S_{im} is 'missing value'.

Table 1. Structure data of online reviews.

Online revies	**CSDs**			
	$\boldsymbol{f_1}$	$\boldsymbol{f_2}$	**…**	$\boldsymbol{f_I}$
	$\boldsymbol{S_1}$	$\boldsymbol{S_2}$	**…**	$\boldsymbol{S_I}$
$\boldsymbol{r_1}$	1	-1	…	Missing value
$\boldsymbol{r_2}$	Missing value	1	…	Missing value
…	…	…	…	…
$\boldsymbol{r_m}$	-1	Missing value	…	-1

3.2 Measuring the effects of customer sentiments toward each CSD on customer satisfaction

GBM (Gradient Boosting Machine) is applied to measure the effects of customer sentiments toward each CSD on rating (customer satisfaction). A boosting model learns data with several simple or weak models and combine them in an adaptive way that leads to a stronger predictor (Cui et al, 2018). The method constructs a set of classifiers like decision trees or regression models. Then it creates a classifier that combines those functions $f(x) = sgn(\sum_i a_i h_i(x)$ that minimize the error. GBM models are generally used in analyzing complex numerical data rather than image or video data. As table 1 shows, all columns consist of -1, 1 and missing value which is an appropriate data set to analyze using GBM. For detail, see Friedman et al. (2001, p. 339).

3.3 Identifying the feature of each CSD from the customer's perspective

3.3.1 Shapley Value

To understand the feature of each CSDs, the author applied Shapley value to interpret the results by GBM. The Shapley value is a solution concept used in game theory which is the average marginal contribution of a feature value across all possible coalitions. Shapley value is meaningful in figuring out how much each CSDs has an influence on customer satisfaction and whether each CSDs have an effect on customer satisfaction in positive way or negative way. Game theory is the study of mathematical models of strategic interaction among rational decision-makers. In terms of game theory, game players are all customer sentiments regarding a topic. Then each player tries to contribute to build customer satisfaction with high or low proportion and with a positive or negative direction. If one review represents the CSD 'a' with positive response and the CSD 'b' with negative response, then two players join in the game whereas other CSD players give up involving the game.

To specify this concept into interpretative machine learning, the author followed Lundberg (2019). It was recently noted that many current methods for interpreting individual machine learning model predictions fall into the class of additive feature attribution models (Lundberg, 2017). This class covers methods that explain a model's output as a sum of real values attributed to each input feature.

Definition 1. Additive feature attribution methods have an explanation model g that is ar linear function of binary variables:

$g(z') = \emptyset_0 + \sum_{i=1}^{M} \emptyset_i z'_i$ where z' $\in \{0, 1\}^M$, M is the number of simplified input features, and $\phi_i \in$ R.

The z'_i variables typically represent a feature being observed (z'_i = 1) or unknown (z'_i = 0) , and $\emptyset_i$'s are the feature attribution values.

As previously described in Lundberg and Lee (2017), an important property of the class of a dditive feature attribution methods is that there is a single unique solution in this class with three d esirable properties: local accuracy, missingness, and consistency. Local accuracy states that the sum of the feature attributions is equal to the output of the function we are seeking to explain. Missingness states the features that are already missing (such that z'_i = 0) are attributed no importance. Consist ency states that changing a model so a feature has a larger impact on the model will never decreas e the attribution assigned to that feature.

Note that in order to evaluate the effect missing features have on a model f, it is necessary to define a mapping h_x that maps between a binary pattern of missing features represented h_x that maps between a binary pattern of missing features represented by z' and the original function input space. Given such a mapping, we can evaluate f(h_x(z')) and so calculate the effect of observing or not observing a feature (by setting z_i' = 1 or z_i' = 0).

To compute Shapley values we define $f_x(S) = f(h_x(z')) = E[f(x)|x_S]$ where S is the set of non-zero indexes in z' (Figure 2), and $E[f(x)|x_S]$ is the expected value of the function on a subset of the input features. Shapley values combine these conditional expectations with the classic Shapley values from game theory to attribute ϕ_i values to each feature:

$$\phi_i = \sum_{S \subseteq N/\{i\}} \frac{|S|!(M-|S|-1)!}{M!} [f(x_{S\cup\{i\}}) - f(x_S)], \text{ when N is the set of all input features}$$

Figure 3.

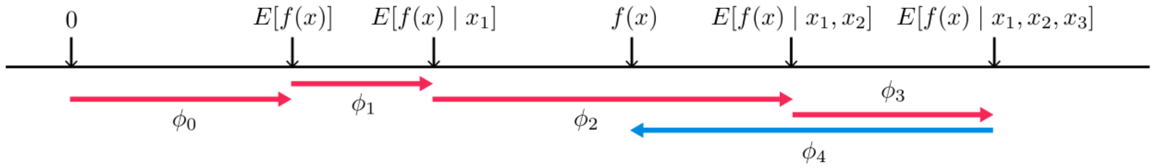


Figure 2: Shapley values explains the output of a function f as a sum of the effects ϕ_i of each feature being introduced into a conditional expectation. Importantly, for non-linear functions the order in which features are introduced matters. Shapley values result from averaging over all possible orderings.

4. Empirical study

In this section, an empirical study is conducted to demonstrate the proposed method. The empirical study measures customer satisfaction of smartwatch products with various brands and product lines. The author used an open source distribution of the Python language and data analysis packages such as Keras, Scikit learn, and Tensorflow.

4.1 Experimental data

As mentioned above, the author selected smartwatch as a target product. The author collected 40,000 online reviews to use in the study from Amazon.com, the largest e-commerce platform in the worldwide. The collected data contain 5 major brands with 13 product lines (Table 2).

Table 2. The details of the experimental data.

Brand	Model	# of Review
Apple	Series 3, Series SE, Series 6	5080
Fitbit	Fitbit Versa 2, Fitbit Versa 3	5870
Letsfit	Letsfit	5120
Samsung	Galaxy Watch 1, Galaxy Watch Active 2, Galaxy Watch 3, Gear, Galaxy Fit, Galaxy Fit 2	17220
Willful	Willful	5000

4.2 Mining customers' sentiments toward CSDs from online reviews.

As mentioned in section 3.1.1, the LDA was applied to extract CSDs from 38,000 online reviews. As results of coherence test and perplexity test, the optimal number of CSDs was determined as 13. In preprocessing reviews for LDA, reviews were split into vocabularies. As the purpose of conducting is extracting CSDs, emotional expressions, stopwords, and brand or product related words are unnecessary in extracting CSDs. Hence, they are excluded from the LDA analysis. Then, lemmatization was implemented to standardize words. After that, the dictionary for LDA was composed and each word was numbered. Through the LDA model, 13 topics containing 20 related vocabularies and their α values are found (Table 3). And the author named each topic according to their vocabularies (Table 4). These topic names are regarded as smartwatch CSDs.

Table 3. Topic words and α value

Topic 1	α value	Topic 2	α value	…	Topic 13	α value
band	0.052	message	0.038	…	sleep	0.057
wrist	0.031	work	0.037	…	step	0.045
easy	0.029	notification	0.029	…	heart	0.043
face	0.019	see	0.013	…	easy	0.041
look	0.018	iphone	0.012	…	rate	0.040
screen	0.018	email	0.012	…	track	0.039
color	0.018	answer	0.010	…	tracking	0.038
use	0.015	alert	0.009	…	use	0.030
…	…	…	…	…	…	…

α value – a distribution of the specific word in the topic

Table 4. Smartwatch CSDs

CSD 1	CSD 2	CSD 3	CSD 4	CSD 5	CSD 6	CSD 7
Band	Connection	Appearance	Price	Time	Display	Sport
CSD 8	**CSD 9**	**CSD 10**	**CSD 11**	**CSD 12**	**CSD 13**	
Warranty	E-commerce	Software	Waterproof	Battery	Sleep	

According to the process of assigning CSDs in reviews, the author labeled a CSD per one sentence in a review based on $\underset{CSD\ X}{\operatorname{argmax}} \sum \alpha$. In detail, in one sentence, the sums of α values per topic are calculated and the sentence is labeled as a specific CSD that has the largest α. If all α values are 0, which implies none of the topic related words are not included in the sentence, then that sentence is excluded in the analysis. An example of assigning CSDs is demonstrated below (Table 5).

Table 5. Topic Index

	topic_index	topic_value
0	-1	0.000
1	6	0.076
2	-1	0.000
3	11	0.056
4	-1	0.000
...	...	...
211257	9	0.067
211258	10	0.023
211259	11	0.108

Concerning the process of identifying the sentiment orientations of each CSD in the reviews of smartwatch, a Transformer classifier is trained using 144,179 labelled online reviews. Then, using Transformer classifier, the sentiment orientation of each CSD in each online review is determined.

4.3 Measuring the effects of customer sentiments toward CSDs on customer satisfaction

The author transformed data coding suitably to figure out the effect of customer sentiments toward CSDs on customer satisfaction (section 3.1.2.). Then, GBM model was trained to measure the influences of customer sentiments toward CSDs on customer satisfaction. The hyperparameters of XGB are set as: Number of estimators = 10000, Learning rate = 0.05, Max depth = 8, Column samples per tree = 0.4, Subsample = 0.8, Evaluation Metric = RMSE (Root Mean Squared Error). Stratified k-Fold cross-validation was conducted to have reliable results (n = 10, Shuffle sampling on each iteration).

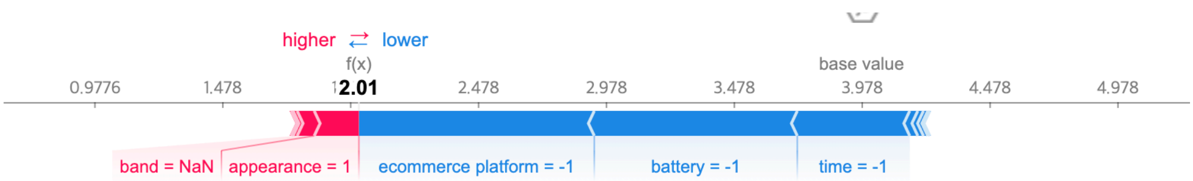
Table 6. Data Coding

	band	connection	appearance	price	time	display	sport	warranty&issue	ecommerce platform	app&software	waterproof&damage	battery	track&sleep
0	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	1.0	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	1.0	1.0	1.0	NaN	NaN	1.0	NaN	NaN	NaN	NaN
2	NaN	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	-1.0	NaN	NaN
3	NaN	NaN	1.0	NaN	NaN	-1.0	NaN	1.0	-1.0	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN	1.0	NaN	NaN	NaN	1.0	1.0	1.0	1.0	1.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
33919	NaN	1.0	NaN	NaN	1.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
33920	NaN	NaN	NaN	NaN	NaN	NaN	1.0	NaN	NaN	NaN	NaN	NaN	NaN
33921	NaN	NaN	NaN	NaN	NaN	NaN	1.0	1.0	1.0	NaN	NaN	NaN	NaN
33922	NaN	NaN	NaN	NaN	NaN	NaN	1.0	1.0	NaN	1.0	1.0	1.0	1.0
33923	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

4.5 Interpreting the result using Shapley value

To classify CSDs based on Kano model, the author analyzed each CSD's feature by applying Shapley value concept. SHAP package which supports interpretative machine learning tool was used in this subsection. First, two plots explain how to get from the base value E$[f(z)]$ that would be predicted if we did not any features to the current output $f(x)$. The plots show the ordering. When the model is non-linear or the input features are not independent, however, the order in which features are added to the expectation matters, and the SHAP values arise from averaging the $\emptyset_i$ values across all possible orderings. As plot illustrates, one of the reviews has the positive response on appearance, the negative response on ecommerce platform, the negative response on battery, and the negative response. Negative sentiments on CSDs have negative effects on customer satisfaction, whereas the positive sentiment toward appearance slightly influences on customer satisfaction. In addition, no response toward band also positively effects on customer satisfaction. And the negative sentiment toward ecommerce platform has the largest impact on customer satisfaction.

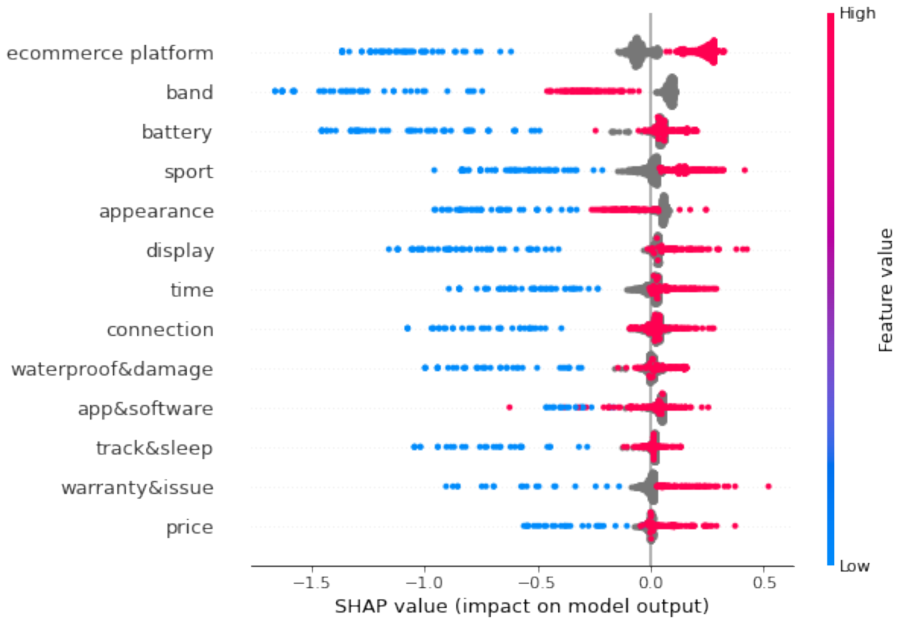
Figure 4. Force plot of one review



After we understand the process to get to the predicted value of one review, we need to see the range and distribution of impacts that feature has on the model's output, and how the feature's value relates to its' impact. SHAP summary plot leverage individualized feature attributions to convey all these aspects of a feature's importance while remaining visually concise. Features are first sorted by their global impact $\sum_{j=1}^{N} \emptyset_i^{(j)}$, then dots representing the SHAP values $\emptyset_i^{(j)}$ are plotted horizontally, stacking vertically when they run out of space.

The SHAP summary plot in Figure 5 represents 13 features GBM customer satisfaction model. The higher the SHAP value of a feature, the higher customer satisfaction score can be gotten. Every customer in the dataset is run through the model and a dot is created for each feature attribution value, so one person gets one dot on each feature's line (positive response: red dot, negative response: blue dot, no response: gray dot). Summary plot arranged CSDs in order of their feature importance (impact power on DV) which is the average of absolute Shapley values.

Figure 5. SHAP summary plot



As the summary plot demonstrates, Shapley values of negative sentiments on CSDs are larger than those of positive sentiments on CSDs. This is because the base value starts with 3.987. Since the maximum customer satisfaction value is 5, Shapley values can't be large enough. Negative effects of e-commerce platform, band, and battery CSDs generally contain large negative Shapley value. This implies that if customers are not satisfied with those features, they heavily give negative effects on customer satisfaction. In addition, all or some of positive sentiments toward band, appearance, waterproof & damage, and track & sleep feature have a negative Shapley value, which means that the customers are unsatisfied on smartwatches due to their positive sentiment on a product. It is surprising that ecommerce platform has the largest feature importance. This is due to the characteristic of online review. Generally, customers write their review within in a month in which their satisfaction are highly influenced by buying process. On the other hands, price get the lowest feature importance. Online reviews are written after they buy a smartwatch, which means they accept a product price. Hence, it can't be an important factor.

4.6 Identifying the category of each CSD from the customer's perspective

According to Figure 5, the category of each CSD of smartphones can be identified. The CSDs are identified as excitement CSDs, must-be CSDs, performance CSDs, reverse CSDs, and indifferent CSDs. Table 7 demonstrates the expected Kano model category by the author and identified Kano model category. As Table 7 shows, e-commerce platform, band, appearance, connection, app & software, and tack & sleep CSD are recognized different to expected Kano model category.

Table 7. Identified Kano model category

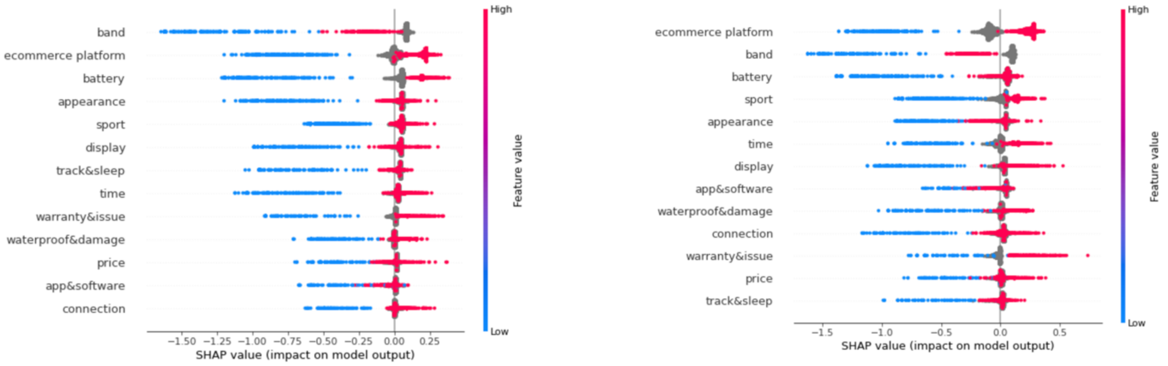
Customer Satisfaction Dimension (CSD)	Expected Kano model category	Analyzed Kano model category
e-commerce platform	Must-be	Performance
band	Performance	Must-be
battery	Performance	Performance
sport	Performance	Performance
appearance	Performance	Must-be
display	Performance	Performance
time	Performance	Performance
connection	Attractive	Performance
waterproof & damage	Must-be	Must-be
app & software	Attractive	Must-be & Reverse
track & sleep	Performance	Must-be
warranty & issue	Performance	Performance
price	Performance	Performance

4.7 Comparison between brands (Samsung and Apple)

In this part, the characteristics of brands are compared using their own summary plot. First, two brands of feature importance is somewhat different. Apple users are most sensitive at band CSDs and more considerable at appearance than Samsung users. On the other hands, Samsung users are more focusing on time, app & software, and connection CSDs than Apple users. Second, Apple's battery CSD is identified as performance CSD whereas Samsung's battery CSD is classified as must-be CSD. Not only that, Samsung's appearance CSD is much close to must-be CSD than Apple's appearance CSD. Last but not least, display CSD of Samsung can be categorized to performance CSD, but display CSD of Apple is between performance CSD and must-be CSD.

Figure 6. Summary plot of Apple products

Figure 7. Summary plot of Samsung products

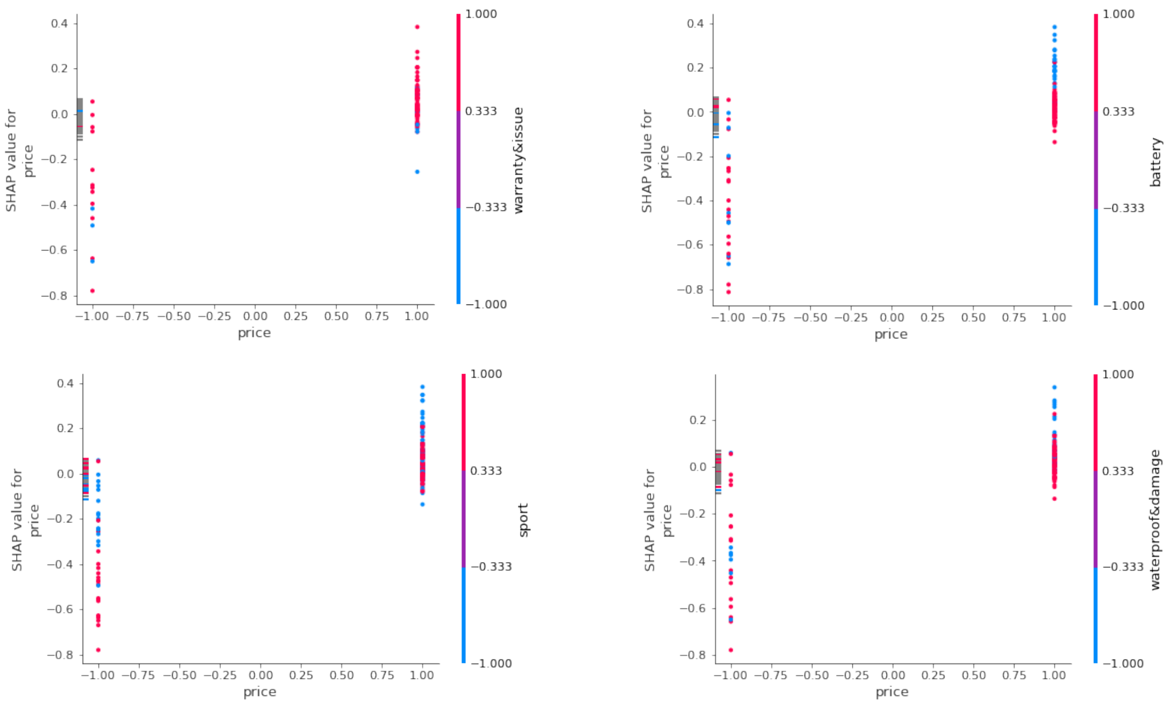


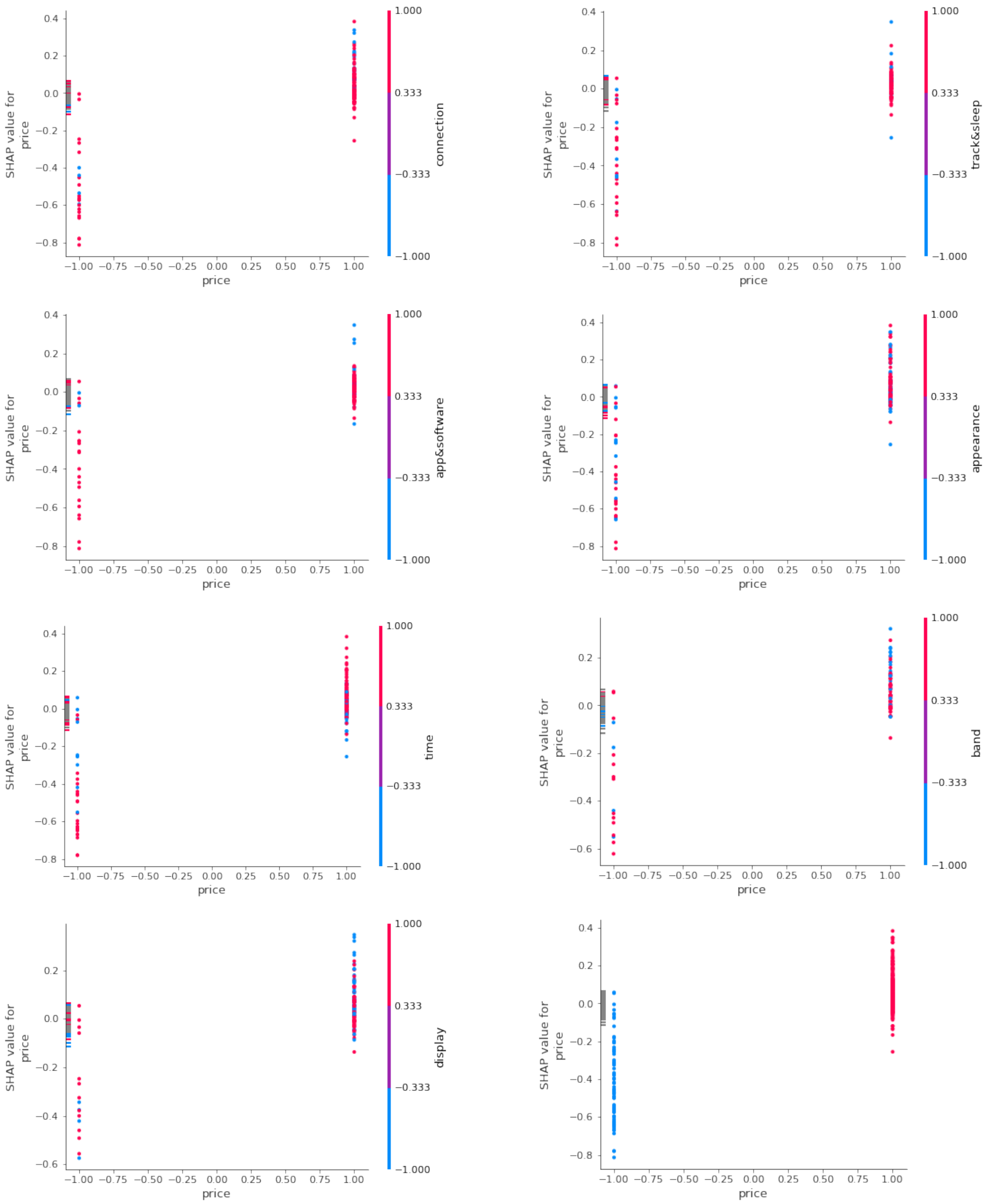
4.8 Interaction Effects between CSDs

In this subsection, the author tried to interpret interaction effects between price CSDs and other CSDs. Because sentiment orientations toward price are typically depends on other CSDs. Dependence plots represent the expected output of a model when the value of a specific variable (or group of variables) is fixed. The values of the fixed variables are varied and the resulting expected model output is plotted. Plotting how the expected output of a function changes as we change a feature helps explain how the model depends on that feature.

As battery and sport dependence plot illustrates, Shapley values of price CSD is high even though customers evaluated smartwatch's battery or sport function negatively. However, Shapley values of price CSD is also high when warranty & issue CSD and time CSD got positive evaluation from users. Moreover, negative Shapley values of price CSD are shown although customers are positive on other CSDs at the same time. This implies negative sentiment orientation toward price is independent to other CSDs.

Figure 8. Price Dependence Plot





5. Conclusion

5.1 Contribution

The major contributions of this paper can be summarized as follows.

First, this study proposes a novel method for modelling customer satisfaction from online reviews. Although the method modified Bi et al (2019) method, big data analytic technique used in this model are much reliable and valid than existing study. Also, as this method applied Shapley value which is validated and accepted from economic field, the method can interpret the result and find implications more deeply.

Secondly, since the method analyzed through online reviews, the method enables to reflect VoC (Voice of Customer) directly to the results.

Thirdly, the number of sample data was larger than pre studies, which means the result includes a large number of customer voice.

Fourth, this study not only analyzed smartwatch products in comprehensive view, but also investigated and compared brands and product lines.

Fifth, the study identified CSDs feature based on Kano model, which enables to understand features of CSDs and give insight on product improvement or development.

Sixth, through the study, the order of feature importance and interaction effects among CSDs are identified and that managers can improve their operations, marketing, R&D strategies more efficiently and effectively.

5.2 Limitation

However, the study also has some limitations in which the author recommends to future researchers.

First, this study only classified sentiment orientations in a binary way (positive or negative). However, customer responses are more complex that can't be understood in this classification way. It will be better to classify sentiment orientation according to power (degree) of sentiment or a type of sentiment (happy, delight, sad, mad and so on).

Second, it is required to collect more data with more various brands and product lines. This study only considers 5 brands, 13 product lines, which means a variety of smartwatches aren't analyzed. If future research contains those product lines, then the result would be much representative.

Third, studies concerning improving or developing product based on the proposed method are recommended. A purpose of measuring and understanding customer satisfaction is to add positive customer value to a product. Thus, the managerial framework should be extended to the product improvement stage.

6. Reference

- Amirata Ghorbani and James Zou. 2019. "Data Shapley: Equitable Valuation of Data for Machine Learning." Proceedings of the 36th International Conference on Machine Learning, PMLR 97:2242-2251.
- Anderson, E. W., C. Fornell, and D. R. Lehmann. 1994. "Customer Satisfaction, Market Share, and Profitability: Findings from Sweden." Journal of Marketing 58 (3): 53–66.
- Blei, D. M., A. Y. Ng, and M. I. Jordan. 2003. "Latent Dirichlet Allocation." Journal of Machine Learning Research 3: 993–1022. Brown, A., and J. J. Reade. 2019. "The Wisdom of Amateur Crowds: Evidence from an Online Community of Sports Tipsters." European Journal of Operational Research 272 (3): 1073–1081.
- Cambria, E., D. Das, S. Bandyopadhyay, and A. Feraco. 2017. A Practical Guide to Sentiment Analysis. Heidelberg: Springer. Chaturvedi, I., E. Ragusa, P. Gastaldo, R. Zunino, and E. Cambria. 2017. "Bayesian Network Based Extreme Learning Machine for Subjectivity Detection." Journal of the Franklin Institute 355 (4): 1780–1797.
- Chiang, D. M., R. S. Guo, and F. Y. Pai. 2008. "Improved Customer Satisfaction with a Hybrid Dispatching Rule in Semiconductor Back-end Factories." International Journal of Production Research 46 (17): 4903–4923.
- Chong, A. Y. L., E. Ch'ng, M. J. Liu, and B. Li. 2017. "Predicting Consumer Product Demands via Big Data: the Roles of Online Promotional Marketing and Online Reviews." International Journal of Production Research 55 (17): 5142–5156.
- Cui, R., S. Gallino, A. Moreno, and D. J. Zhang. 2018. "The Operational Value of Social Media Information." Production and Operations Management 27 (10): 1749–1769.
- Culotta, A., and J. Cutler. 2016. "Mining Brand Perceptions from Twitter Social Networks." Marketing Science 35 (3): 343–362. Decker, R., and M. Trusov. 2010. "Estimating Aggregate Consumer Preferences from Online Product Reviews." International Journal of Research in Marketing 27 (4): 293–307.
- Farhadloo, M., R. A. Patterson, and E. Rolland. 2016. "Modeling Customer Satisfaction from Unstructured Data Using a Bayesian Approach." Decision Support Systems 90: 1–11.
- Farhadloo, M., and E. Rolland. 2013. "Multi-class Sentiment Analysis with Clustering and Score Representation." IEEE 13th International Conference on Data Mining Workshops (ICDMW), TX, USA, 904–912.
- Grewal, R., J. A. Cote, and H. Baumgartner. 2004. "Multicollinearity and Measurement Error in Structural Equation Models: Implications for Theory Testing." Marketing Science 23 (4): 519–529.
- Guo, Y., S. J. Barnes, and Q. Jia. 2017. "Mining Meaning from Online Ratings and Reviews: Tourist Satisfaction Analysis Using Latent

Dirichlet Allocation.” Tourism Management 59: 467–483.
- Hallowell, R. 1996. “The Relationships of Customer Satisfaction, Customer Loyalty, and Profitability: An Empirical Study.” International Journal of Service Industry Management 7 (4): 27–42.
- Hui, T. K., D. Wan, and A. Ho. 2007. “Tourists’ Satisfaction, Recommendation and Revisiting Singapore.” Tourism Management 28 (4): 965–975.
- Kano, N., N. Seraku, F. Takahashi, and S. Tsuji. 1984. “Attractive Quality and Must-be Quality.” Journal of Japanese Society for Quality Control 14 (2): 39–48.
- Kim, W. G., C. Y. N. Ng, and Y. S. Kim. 2009. “Influence of Institutional DINESERV on Customer Satisfaction, Return Intention, and Word-of-mouth.” International Journal of Hospitality Management 28 (1): 10–17.
- Li, Y. L., Y. F. Du, and K. S. Chin. 2018. “Determining the Importance Ratings of Customer Requirements in Quality Function Deployment Based on Interval Linguistic Information.” International Journal of Production Research, doi:10.1080/00207543.2017.1417650.
- Liu, Y., J. W. Bi, and Z. P. Fan. 2017a. “Ranking Products Through Online Reviews: A Method Based on Sentiment Analysis Technique and Intuitionistic Fuzzy set Theory.” Information Fusion 36: 149–161.
- Liu, Y., J. W. Bi, and Z. P. Fan. 2017b. “Multi-class Sentiment Classification: The Experimental Comparisons of Feature Selection and Machine Learning Algorithms.” Expert Systems with Applications 80: 323–339.
- Liu, Y., L. Wang, X. V. Wang, X. Xu, and L. Zhang. 2018. “Scheduling in Cloud Manufacturing: State-of-the-art and Research Challenges.” International Journal of Production Research, doi:10.1080/00207543.2018.1449978.
- Oliver, R. L. 1980. “A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions.” Journal of Marketing Research 17 (4): 460–469.
- Qi, J., Z. Zhang, S. Jeon, and Y. Zhou. 2016. “Mining Customer Requirements from Online Reviews: A Product Improvement Perspective.” Information & Management 53 (8): 951–963.
- Scott M. Lundberg, Gabriel G. Erion, and Su-In Lee. 2019, “Consistent Individualized Feature Attribution for Tree Ensembles.”, Cornell University, Preprint at http://arxiv.org/abs/1802.03888.
- Sepehr, S., and M. Head. 2018. “Understanding the Role of Competition in Video Gameplay Satisfaction.” Information & Management 55 (4): 407–421.
- Shapley, L. S. 1998. “The Shapley Value: Essays in honor of Lloyd S. Shapley.”, Cambridge University Press.
- Tirunillai, S., and G. J. Tellis. 2014. “Mining Marketing Meaning from Online Chatter: Strategic Brand Analysis of big Data Using Latent Dirichlet Allocation.” Journal of Marketing Research 51 (4): 463–479.
- Woodruff, R. B., E. R. Cadotte, and R. L. Jenkins. 1983. “Modeling Consumer Satisfaction Processes Using Experience-based Norms.” Journal of Marketing Research 20 (3): 296–304.