

ORIGINAL RESEARCH

A study of the possibilities of text mining and machine learning for score evaluation and review content

Yuko Taniguchi*, Kazuhiko Tsuda

Dept. Of Risk and Resilience Engineering, Graduate School of System and Information Engineering, Tsukuba University, Tsukuba, Japan

Received: April 29, 2020

Accepted: June 11, 2020

Online Published: June 27, 2020

DOI: 10.5430/air.v9n1p1

URL: <https://doi.org/10.5430/air.v9n1p1>

ABSTRACT

With the widespread use of the Internet, there are more and more opportunities to purchase a variety of products through online shopping. The opportunities are not only for small products such as books, but also for home appliances. Previously, when purchasing a product, users who wanted to buy a product would visit a store and get expert advice on what to buy. Now, however, customers consider reviews on the Internet to be more important information for considering the products to be purchased. And evaluation page consists of an overall evaluation, an evaluation of each feature, and comments, which are word of mouth. The overall evaluation and the evaluation of each feature is often a score evaluation, and organized information such as the average and the distribution of scores are presented. However, it is difficult to read all the comments that are word of mouth because they are often enumerated as is. Therefore, in this study, we created a system to label which features people commented on in response to the word of mouth comments using data from the TV's comprehensive evaluation page. 2392 TV evaluation results from Sony.com were used. From the extracted data, text mining was performed on the comments, which are word of mouth, followed by labels of which features are commented on. When 80% of the test data was prepared and implemented against 20% of the learning data, the label was predicted with 77% accuracy. From this study, we used text mining to label the comments, which are customer impression. from the current study, text mining was used to label the comments, which are customer impression. The results and score ratings were used to identify customer trends.

Key Words: Text mining, Word of mouth, Machine learning

1. INTRODUCTION

With the widespread use of the Internet, the use of online shopping at home contains to increase. Online shopping at home is not only for books and clothes, but also for home appliances in on the rise. And it now accounts for 10% of all consumer electronics products. In addition, the number of people buying home appliances online is expected to grow 1.8 times over the next eight years, starting in 2012. In the past, people considering the purchase of home appliances

used to go to the store and ask for advice from clerks and manufacturer's experts in order to select the products they wanted to buy. However, today, many buying prospective buyers are deciding which products to buy by gathering and comparing information on potential purchases that exist online. The information on the product that the purchase collects is not only on the page that introduces the product or describe the specifications, but also on the product evaluation page. It is said that word of mouth on the product evaluation page

***Correspondence:** Yuko Taniguchi; Email: s1830140@s.tsukuba.ac.jp; Address: Dept. Of Risk and Resilience Engineering, Graduate School of System and Information Engineering, Tsukuba University, Tsukuba, Japan.

is useful information when considering the purchase of a product, and many people who are considering the purchase of a product use it.

There is a product evaluation page written by the users of the product, which mainly consists of a score evaluation and word of mouth information. In a score evaluation, information is often organized, such as averages and the distribution of scores. On the other hand, word of mouth information is often just enumerated as it is. In the case of a product that is selling a lot, there is a lot of word of mouth information, and it is difficult for a product purchaser to decipher all the word of mouth information on the site where the word of mouth is collected. However, when looking at the word of mouth information written by users, there are many comments about how easy it is to use the product and what they noticed after the purchase that they didn't notice in the store or before the purchase, which often gives them a hint to purchase.

2. PROBLEM AWARENESS AND BACKGROUND

2.1 Commoditization and business stalemate

Many companies try to make products that are more attractive or more technologically superior than their competitors in order to differentiate themselves from other companies. As a result, there are many cases in which consumer electronics manufacturers are unable to turn a business into a profit due to inflated costs.

In addition, the company has become a one-sided manufacturer of products, has not communicated the attractiveness of its products to costumers properly, and has failed to understand how customers feel, and has made investments that do not meet customers' needs, resulting in a dead end for the business.

Therefore, we believe it is important to correctly collect what customers are thinking and feeling, and then find hints for marketing approaches and product improvements.

2.2 The spread of the Internet and product purchases by word-of-mouth

With the spread of the Internet,^[1] the number of Internet users and the penetration rate of the population in Japan has been on an upward trend year by year; when comparing 2014 and 2015, the number of Internet users increased by 280,000 per year, and the penetration rate as a percentage of the population increased by 0.2%. The increase in the number of users is not a phenomenon that has occurred only in Japan. A study of the global Internet shows,^[2,3] similarly, the number of Internet users is increasing year by year, with an estimated 2.923 billion people using the Internet in 2014, compared to only 1.024 billion people in 2005. The average one-year growth rate is 12.4%, which is much higher than the increase in the number of Internet users in Japan. The results of the latest global Internet survey are shown in Figure 1. In 2019, there will be an estimated 5.36 billion people.

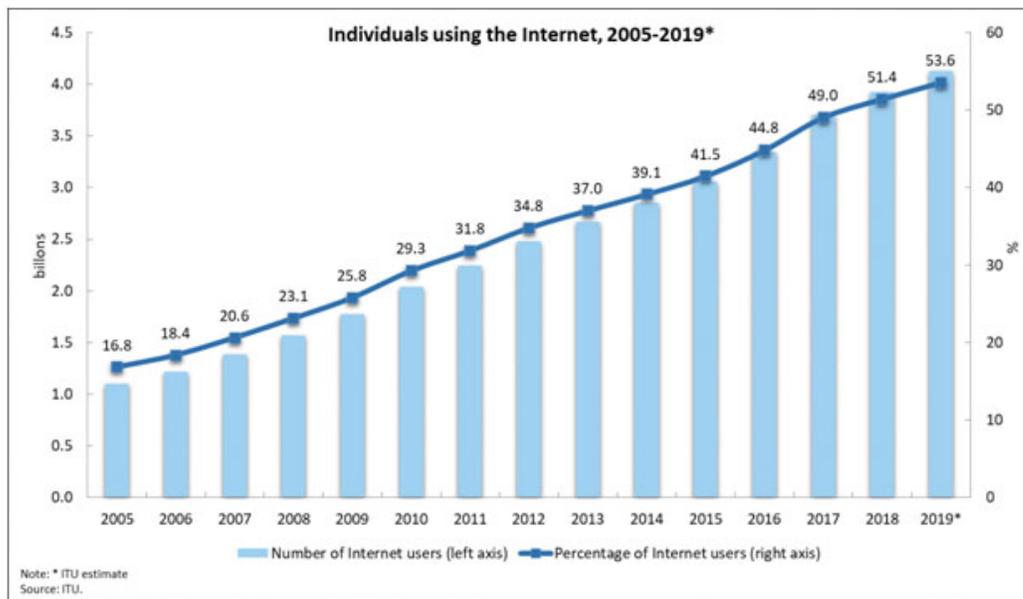


Figure 1. Individuals using the Internet, 2005-2019

The widespread use of the Internet has led to the creation and dissemination of various services through the Internet. One of them is online shopping. According to the Household Consumption Survey,^[4] the percentage of households that ordered through the Internet was 27.8% in 2016, compared to only 5.3% in 2002, a 5.3-fold increase in 16 years. In addition, many people cited the advantages of online shopping, such as the ability to shop without visiting a real store, the ability to shop 24 hours a day, and low prices. In the past, people considering the purchase of home appliances used to go to a store and ask for advice from a clerk or expert in order to select the product they wanted to buy. Today, however, many prospective purchasers are less likely to visit a store. Therefore, the purchaser must judge the product to be purchased by himself or herself and decides the product to be purchased by collecting and comparing the information of the candidate products that exist on the Internet.

It is possible to obtain information such as specifications from the catalog information provided by the manufacture. However, since information such as usability cannot be obtained from catalogs, it has become necessary from product purchasers to obtain information that cannot be obtained from catalogs by word of mouth.

2.3 Changes in support

With the spread of the Internet, the awareness of product support has also changed. In the past, the call center was the only touch point with the customer in the support system. However, there are now a wide variety of touch points with customers, such as providing information through online manuals and opinions and inquiries about products through social media.^[5,6] We have also found that customer attitudes have changed, and if they receive good support of social media, they are more likely to choose the same manufacturers' products next time, and to extend their experience through social media.

2.4 Precious research on word of mouth

Nowadays, there are many studies on word of mouth information.

As a way to quantify product features from word of mouth information, there was one about mirrorless cameras from the Japanese site, kakaku.com.^[7] Hashimoto proposed a method to quantitatively understand the added value proposed by the manufacturers and the response of users for seven major camera manufactures in Japan and explained the difference between them. It is understood that each manufacturer forms

the value of its products based on user needs, while Sony forms the value of getting users to accept its proposals and appeals. In addition, one of them examined the use of Twitter as a way to establish brand management using word of mouth on the Internet and conduct a sensory analysis of the content of Tweets about the brand and its brand.^[8] Bernard reports that a weekly sensory analysis of the brand showed more than 60% of the results were positive, but he argues that the character limit may have had some effect on Twitter. A study on the relationship between score evaluation, and word of mouth information was conducted on Japanese language accommodation reviews.^[9] Tsujii explains that when a customer considering a stay makes a reservation through an Internet reservation site, the reviews posted by the stay users are considered important and suggests a method to extract information that is meaningful to the person considering a stay, and explains the difference by area. There was also an approach regarding Amazon.com as a way to predict the usefulness of EX site reviews.^[10] Soo-Min says that there are more and more reviews in the world, and the quality of these reviews varies, so it is important to find out which ones are useful. As a method, they take a machine learning approach to predict the usefulness of reviews and propose a way to use SVM to display rankings in a real-world dataset collected from Amazon.com.

A study was conducted on the relationship between score ratings and word of mouth information on Japanese product comparison site and accommodation reviews. A proposal for a method of using SVM to display rankings from Amazon's reviews has also been studied. However, none of these previous studies were written in English about the study of score ratings and word of mouth information, and none were written about TV. Also, nothing was written about data cleansing in the methods that use machine learning.

3. REVIEW INFORM AND TEXT MINING

3.1 Review information for product purchase sites

Many European and American consumer electronics manufacturers have product review information on their product purchase pages. The score ratings on the product purchase page are often assigned 5 or 4. It is a natural evaluation in a sense, since it is a person who liked and purchased the product.

As shown in Figure 2, in addition to the overall evaluation, the score evaluation of the product review information often includes evaluation items for product features such as product functions and usability.



Figure 2. Reviews in Home Appliances

In this study, we focus not only on the overall evaluation, but also on word of mouth of product reviews, and propose a method of classifying in the written content by machine learning, without the need to read all of the large amount of review data, by preparing training data.

3.2 Labeling with text mining and machine learning

We used Rapid Miner Studio ver9.6 for text mining and machine learning for labeling. Text mining was carried out in combination with Text Processing, an extension of Rapid Miner Studio.

3.2.1 Data preparation

The step of converting the original data into manageable data for analysis is shown in Figure 3.

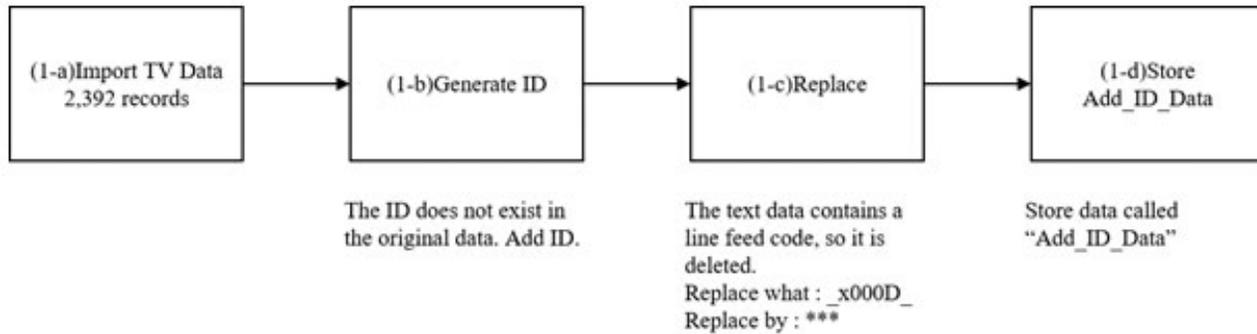


Figure 3. Steps for Data Preparation

We prepared the original data in 1-a and added ID in 1-b. This makes it easier to merge the data later, even if there is no ID in the original data. Then we deleted the line feed code in the sentence in 1-c and saved it as new data add_id_data in 1-d.

was created with 3.642 words registered. Another dictionary was created to remove words that were not needed for the analysis. For example, it consists of model names, proper nouns, numbers, dates, and words that are not necessary for this analysis. A total of 3,401 words was registered in the dictionary.

3.2.2 Prepare dictionary

Two types of dictionaries were used in text mining. A dictionary for correcting notational distortions and a dictionary for correcting unnecessary deletions.

A dictionary for correcting notational distortions is, for example, a dictionary for combining “talk”, “talks”, “talked”, and “talking” into “talk” as shown in Figure 4. A dictionary

¥btold¥b	talk
¥btalks¥b	talk
¥btalking¥b	talk
¥btalked¥b	talk

Figure 4. Prepare dictionary

3.2.3 Implementing text mining

The text mining method for text is shown in Figure 5.

Preparations 3-a data with an ID, Next, since the title and review comments are text, we merged them into one column in 3-b; in 3-c, we selected the ID and merged text that we needed to do text mining; in 3-d, we did the conversion to

text; we had a dictionary d-1 that corrects the notational distortions we created in 3.2.2.; in 3-e, we used the dictionary to convert the text. This will correct any notational distortions in the text. Then, we performed text mining in 3-f, merged the results of 3-g text mining into the original data, and saved that data as clean_add_id_data in 3-h.

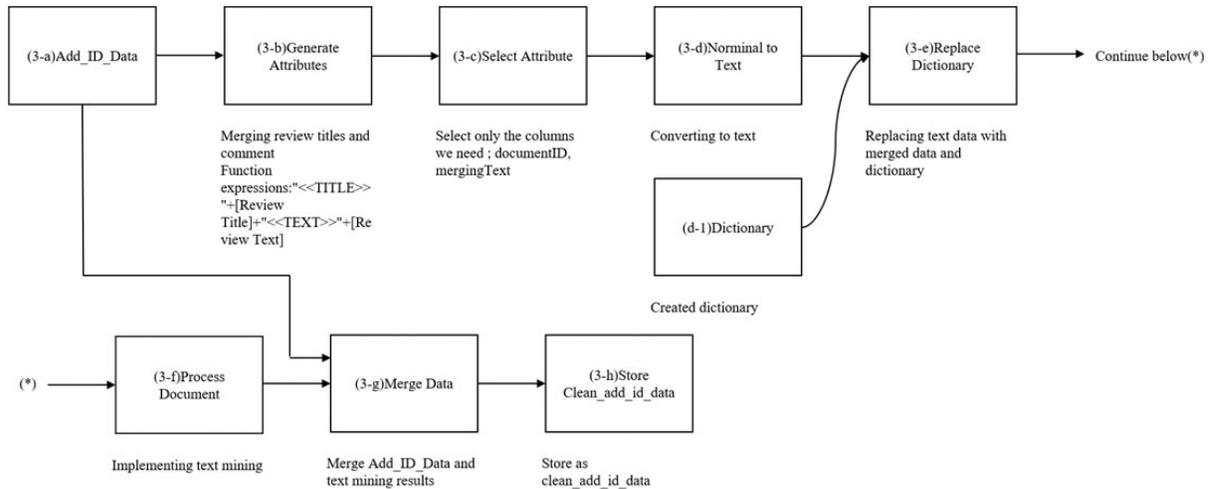


Figure 5. Steps for Implementing text mining

3.2.4 Details of the Process Document

The details of the 3-f process document in Figure 5 are shown in Figure 6. Allowing text mining to pick up only the words

we really need a text mining can be achieved by combining several operators.

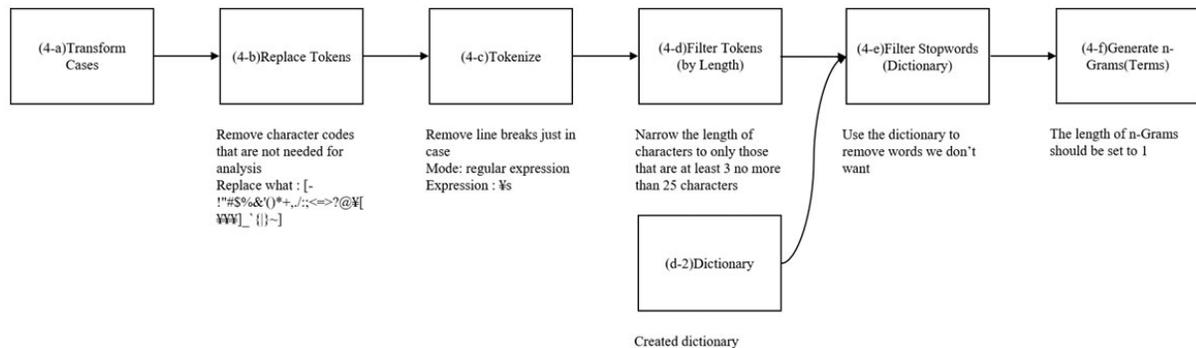


Figure 6. Steps for Details of the Process Document

In 4-a, we converted all the text to lowercase, in 4-b, we removed unnecessary character codes. And in 4-c, we removed line feeds just in case. Then, the length of one character required for analysis in 4-d was set to 3 25 characters. By doing so, short words such as “a” and “it” will be excluded as unnecessary words. Using this dictionary, we asked them

to apply it to the text in 4-e and removed all the unnecessary words for analysis. Finally, we set up the n-Gram in 4-f.

3.2.5 Labeling based on the text

The text mining results are used for labeling. Labeling is done using the filter function, and if a specific word is present, it is set to be labeled. Each step is shown in Figure 7.

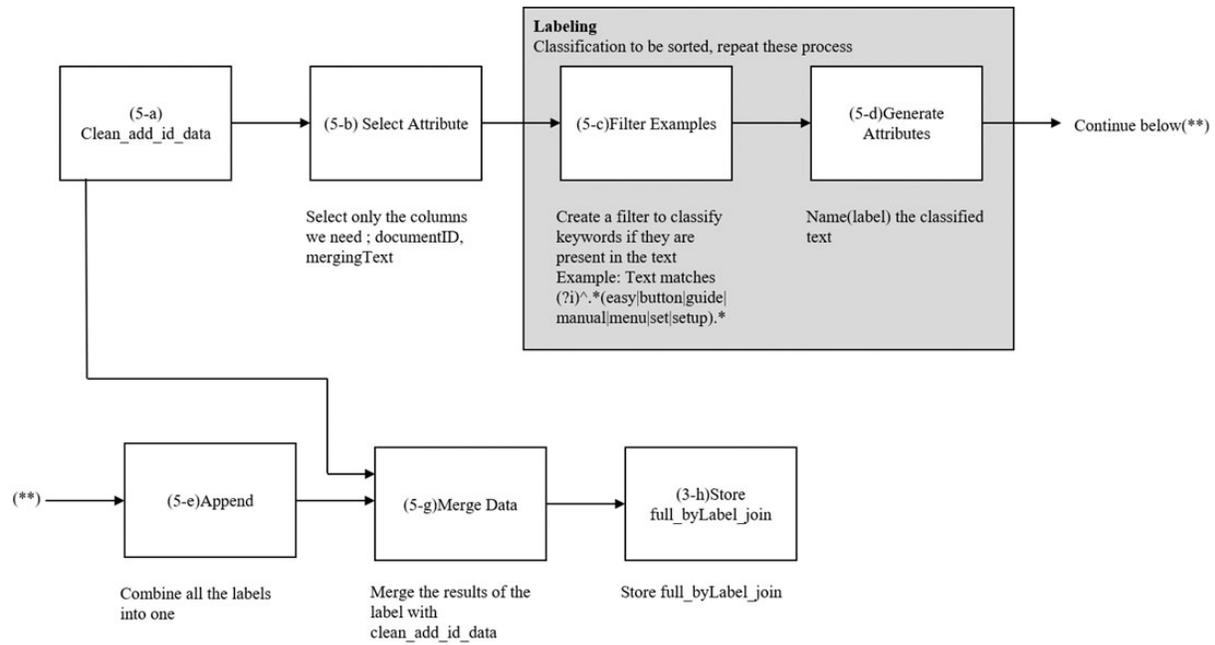


Figure 7. Steps for Labeling based on text

We took the data that were cleaned up in 5-a. Next, we created 5-b set of just the text needed to label it and the IDs needed to merge it later. We set the labels in 5-c and 5-d. Setting up a label requires repeating 5-c and 5-d process, depending on the type of label we want to set up. Using the filter function, we checked the text for the presence of specific words in 5-c and labeled the text that contained specific words in 5-d. In this labeling, when extracting labels from word of mouth information, the labels are based on the help guide rather than sentiment information such as positive and negative. For example, if we wanted to label it “Easy Use”, we set the filter to seven easy of use related words: easy, button, guide, manual menu, set and setup. Of the

seven keywords in “Easy Use”, easy is taken from Easy Use, and the remaining six are taken from the keywords listed in Basic Operations in the Getting Started help guide on Sony’s website. If we wanted to label it “sound”, we set the filter to ten sounds related words: audio, beat, dolby, sound, soundbar, speaker, surround, sync, voice, and volume. Append the labels with 5-e and merge them with clean_add_id_data and 5-f to make each record labeled. Finally, we saved it as full_byLabel_join in 5-g.

3.2.6 Attribute conversion

Figure 8 shows how to change the attributes of the data before performing machine learning.

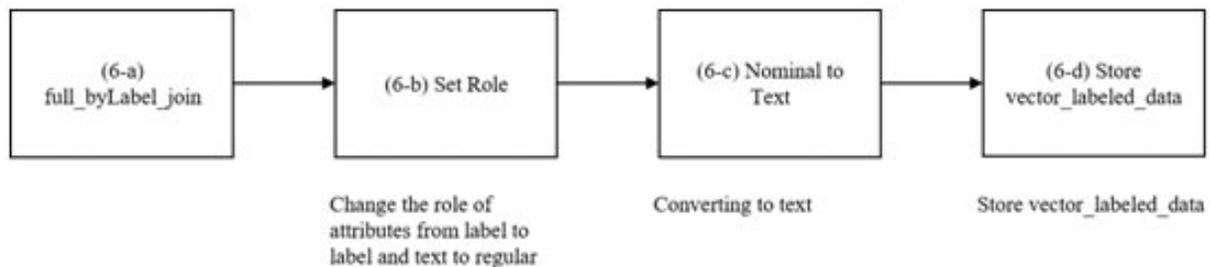


Figure 8. Steps for Attribute conversion

The labeled data were taken in 6-a, the label was used as a label, and the text was used as a regular in 6-b. The text was converted in 6-c, and finally, the data were saved as vector_labeled_data.

3.2.7 Model creation and model application

How the model is created and how the model is applied are shown in Figure 9.

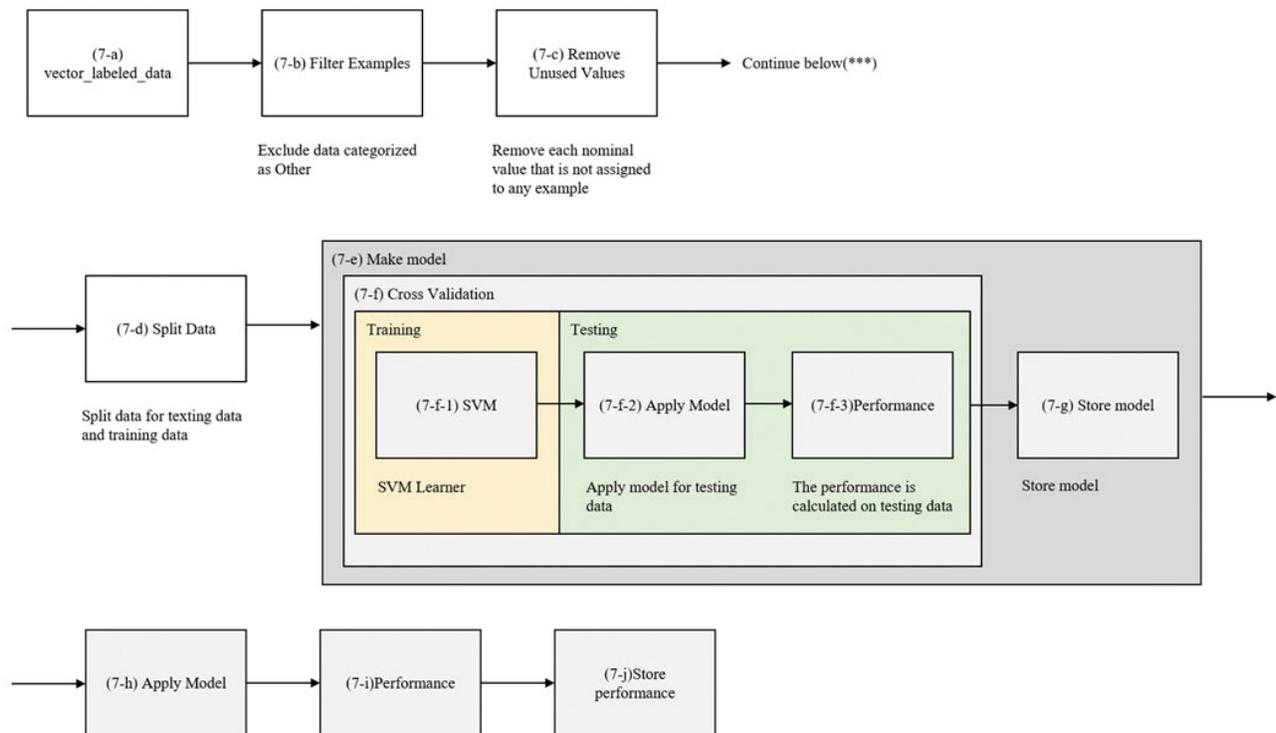


Figure 9. Steps for Model creation and model application

The vector_labeled_data with different attributes were prepared in 7-a, and those assigned to Other were excluded in 7-b. There were 2,392 overall data, of which 358 were for Other, so the base data for creating a non-Other model was 2,034. It is already clean as data, but just in case, we removed each nominal value that is not assigned to any example in 7-c. In order to classify the data into training data and test data, Split Data was used in 7-b to classify the training data into 20% and the test data into 80%.

7-e through 7-g are the steps to create the model. 7-e Make Model includes 7-f Cross Validation and 7-g Store Model. In 7-f Cross Validation, the process is divided into Training Data and Test Data. This time, among several methods, we have been using a support vector machine to create the model. Therefore, 7-f-1 in Training Data is set to SVM Learner. The Apply Model in 7-f-2 fits the model to the Testing Data and confirms the performance when the model is applied in 7-f-3. These processes do not make a lot of sense, but due to the limitations of Rapid Miner Studio, we need to do this. Save the model made by Cross Validation to 7-g.

Then, apply the model we saved with 7-h Apply Model. 7-i Performance examines the performance to check the accuracy of the model applied. Then saved the results with the modeled test data in 7-j Store Performance. By doing these steps, we could create and use the models with Support Vector Machine to label it.

4. SCORE EVALUATION AND WORD OF MOUTH RESULTS

4.1 TV score evaluation results

Table 1 shows the results of the aggregate score that customers gave to the overall evaluation of the TV from Sony’s product site. As can be seen from the results, 66.81% of the total number of respondents gave the highest rating 5, and the total number of respondents who gave the highest rating 4 was 83.19%. On the other hand, only 10.37% of the total respondents received poor scores for 1 and 2. In other words, we can see that the results of the score evaluation are highly biased.

Table 1. Results of Overall Rating

Rating	Count	Rate
5	1,598	66.81%
4	392	16.39%
3	154	6.44%
2	89	3.72%
1	159	6.65%

Next, the evaluation of each function is shown Table 2. Looking at the results of the evaluation of each function, the percentage of people who gave 5 to each function evaluation was 63.55%, compared to 66.81% of the overall evaluation of 5 in Table 1. Looking at the number and percentage of people who gave each function a rating of 5, it can be seen

that Design and Picture Quality have a high percentage of 73.20% and 82.44%, respectively, while Easy Use, Features, and Sound Quality have a low percentage of 58.74%, 61.71%, and 41.64% respectively.

Table 2. Results of Each Function

Rating	Easy Use	Design	Features	PictureQuality	SoundQuality	Average
5	1,405(58.74%)	1,751(73.20%)	1,476(61.71%)	1,972(82.44%)	996(41.64%)	(63.55%)
4	540(22.58%)	379(15.84%)	550(22.99%)	220(9.20%)	731(30.56%)	(20.23%)
3	211(8.82%)	138(5.77%)	180(7.53%)	89(3.72%)	444(18.56%)	(8.88%)
2	112(4.68%)	42(1.76%)	70(2.93%)	37(1.55%)	109(4.56%)	(3.09%)
1	124(5.18%)	82(3.43%)	116(4.85%)	74(3.09%)	112(4.68%)	(4.25%)

Next, Table 3 shows the results of the evaluation of each function for those who scored 5 in the overall evaluation. The results differed from Table 2, but the percentage of Sound Quality rating 5 was low, 54.13%.

Table 3. Results of Each Function with 5 Overall Rating

Rating	Easy Use	Design	Features	PictureQuality	SoundQuality	Average
5	1,252(78.35%)	1,479(92.55%)	1,322(82.73%)	1,564(97.87%)	865(54.13%)	(81.13%)
4	297(18.59%)	114(7.13%)	259(16.21%)	33(2.07%)	479(29.97%)	(14.79%)
3	42(2.63%)	5(0.31%)	13(0.81%)		220(13.77%)	(3.50%)
2	5(0.31%)		3(0.19%)	1(0.06%)	28(1.75%)	(0.46%)
1	2(0.13%)		1(0.06%)		6(0.38%)	(0.11%)

From these results, even those who give the most satisfactory overall rating of 5 are not satisfied with all items. Furthermore, the fact that they gave it a high overall rating even though they were dissatisfied with it means that there may be another point that stand out as satisfying beyond the five items evaluated here.

4.2 Machine learning accuracy

In this case, we used text mining and machine learning to find out what TV shoppers say in their review statements and label them.

The accuracy of the labeling was 77.90% and the accuracy

of each function is shown in Table 4. The labeling of Other, Sound Quality, Design, and Features were all predicted with an accuracy of over 80%. Only Picture Quality, however, had a very low accuracy rate of 36.29%. A closer look at the results shows that the text that should have been categorized as Picture Quality is now categorized as Sound Quality. One method of machine learning, SVM, has good discrimination accuracy even when the dimensions of the data features are large. It also has the advantage that there are fewer parameters to be optimized. However, the SVM were not suitable for the five classifications in this study because it is basically suitable for two classes of classification.^[16-18]

Table 4. Results of Machine Learning Accuracy

	TrueOther	TrueSound	TruePicture	TrueDesign	TrueEasyUse	TrueFeature	Precision
Pred.Other	281	0	4	0	0	0	98.60%
Pred.Sound	0	154	289	0	1	2	34.53%
Pred.Picture	2	0	172	2	7	0	93.99%
Pred.Design	0	11	3	149	4	29	76.02%
Pred.Easy Use	3	9	6	22	540	12	91.22%
Pred.Feature	0	17	0	0	0	195	91.98%
Recall	98.25%	80.63%	36.29%	86.13%	97.83%	81.93%	

The results of the overall evaluation and the labeled results are shown in Table 5. The % result is the value for each rating and item divided by the overall number 2,392. The labeling results used in Table 5 are not results of machine learning,

but rather the results created in Figure 7 when creating the model due to the low accuracy of Picture Quality. Among those who gave 5 to the overall rating, we could see that reviews were given about Easy Use and Picture Quality. The

same trend can be seen when narrowing down to those who gave 5 and 4 to the overall rating. We also found that more than 14% of the total respondents made comments outside of

the five features provided, even though they gave it a rating of 5 or 4.

Table 5. Overall rating and labeled classification results

Rating	Easy Use	Design	Features	PictureQuality	SoundQuality	Other
5	417 (17.43%*)	157(6.56%*)	209(8.74%*)	419(17.52%*)	157(6.56%*)	239(11.36%*)
4	153(6.40%*)	29(1.21%*)	45(1.88%*)	82(3.43%*)	39(1.63%*)	44(2.91%*)
3	59(2.47%*)	11(0.46%*)	17(0.71%*)	27(1.13%*)	16(0.67%*)	24(1.09%*)
2	20(0.84%*)	12(0.50%*)	12(0.50%*)	16(0.67%*)	13(0.54%*)	16(0.61%*)
1	41(1.71%*)	7(0.29%*)	14(0.59%*)	48(2.01%*)	14(0.59%*)	35(1.04%*)

*: The result is the value for each rating and item divided by the overall number 2,392

The results of the item by item evaluation and labeling of Easy Use are shown in Table 6. The results showed that of those who gave Easy Use 5 on their numerical rating, 15.30% mentioned Easy Use in their reviews. From the labeling results, it was found that many users wrote reviews about Picture Quality in addition Easy Use. When the scores of 5 and 4, which are the results of the item by item evaluation, are totaled, 22.03% of the respondents rated Easy Use and 20.99% rated Picture Quality, indicating that Easy Use was written more often. The results of the item by item evaluation and labeling of Sound Quality are shown in Table 8. The

results showed that of those who gave Sound Quality a numerical rating of 5, 10.1% mentioned Sound Quality in their reviews. From the labeling results, it was found that many users wrote reviews about Easy Use in addition to Sound Quality. When the scores for 5 and 4, which are the results of the item by item evaluation, are totaled, 19.73% and 18.69% of the scores for Easy Use and Sound Quality, respectively, indicating that Easy Use was written more often. In other words, it is not possible to find any specific trends in these results when comparing the labeled data with the results of the overall evaluation or the evaluation of each function.

Table 6. Evaluation results of Easy Use and labeled classification results

Rating	Easy Use	Design	Features	PictureQuality	SoundQuality	Other
5	366 (15.30%*)	138(5.77%*)	192(8.03%*)	368(15.38%*)	129(5.39%*)	212(8.86%*)
4	161(6.73%*)	48(2.01%*)	56(2.34%*)	134(5.60%*)	74(3.09%*)	67(2.80%*)
3	76(3.18%*)	18(0.75%*)	24(1.00%*)	47(1.96%*)	14(0.59%*)	32(1.34%*)
2	45(1.88%*)	7(0.29%*)	11(0.46%*)	15(0.63%*)	13(0.54%*)	21(0.88%*)
1	42(1.76%*)	5(0.21%*)	14(0.59%*)	28(1.17%*)	9(0.38%*)	26(1.09%*)

*: The result is the value for each rating and item divided by the overall number 2,392

Table 7 is the same as Table 6, but with a different way of expressing the propositions; in Table 6, each value is divided by the total number, but in Table 8, each value is divided by the total of each function. In other words, at the point where Easy Use and rating 5, the total number labeled Easy Use is 690, and 249/690 is used to arrive at the value. As a result, when rating is 5, there is no significant difference

between the functions. When the percentages of rating 1 and 2 are combined, Easy Use 12.61%, Design 5.56%, Features 8.42%, Picture Quality 7.26%, Sound Quality 9.21%, and Other 13.13%. when Other is excluded, Easy Use is the highest in terms of percentage, indicating that those who are dissatisfied with Easy Use are more likely to make comments on Easy Use.

Table 7. Evaluation results of Sound Quality and labeled classification results

Rating	Easy Use	Design	Features	PictureQuality	SoundQuality	Other
5	246 (10.28%*)	87(3.64%*)	134(5.60%*)	243(10.16%*)	112(4.68%*)	174(7.27%*)
4	226(9.45%*)	64(2.68%*)	94(3.93%*)	204(8.53%*)	58(2.42%*)	85(3.55%*)
3	151(6.31%*)	43(1.80%*)	48(2.01%*)	102(4.26%*)	35(1.46%*)	65(2.72%*)
2	41(1.71%*)	8(0.33%*)	11(0.46%*)	18(0.75%*)	12(0.50%*)	19(0.79%*)
1	26(1.09%*)	14(0.59%*)	10(0.42%*)	25(1.05%*)	22(0.92%*)	15(0.63%*)

*: The result is the value for each rating and item divided by the overall number 2,392

Table 8. Evaluation results of Easy Use and labeled classification results, divided by total of each item

Rating	Easy Use	Design	Features	PictureQuality	SoundQuality	Other
5	336(53.04% ^{**})	138(63.89% ^{**})	192(64.65% ^{**})	368(62.16% ^{**})	129(53.97% ^{**})	212(59.22% ^{**})
4	161(23.33% ^{**})	48(22.22% ^{**})	56(18.86% ^{**})	134(22.64% ^{**})	74(30.96% ^{**})	67(18.72% ^{**})
3	76(11.01% ^{**})	18(8.33% ^{**})	24(8.08% ^{**})	47(7.94% ^{**})	14(5.86% ^{**})	32(8.94% ^{**})
2	45(6.52% ^{**})	7(3.24% ^{**})	11(3.70% ^{**})	15(2.53% ^{**})	13(5.44% ^{**})	21(5.87% ^{**})
1	42(6.09% ^{**})	5(2.31% ^{**})	14(4.71% ^{**})	28(4.71% ^{**})	9(3.77% ^{**})	26(7.26% ^{**})

*: The result is the value for each rating and item divided by total of each item

Similarly, Table 9 shows the percentage of Sound Quality calculated by the total of each item. When the percentage of rating 1 and 2 are combined, Easy Use 9.71%, Design 10.19%, Features 7.07%, Picture Quality 7.26%, Sound Qual-

ity 14.23%, and Other 9.50%. Sound Quality is the highest in terms of percentage, indicating that those who are dissatisfied with Sound Quality are more likely to make comments on Sound Quality.

Table 9. Evaluation results of Sound Quality and labeled classification results, divided by total of each item

Rating	Easy Use	Design	Features	PictureQuality	SoundQuality	Other
5	246 (35.65% ^{**})	87(40.28% ^{**})	134(45.12% ^{**})	243(41.05% ^{**})	112(46.86% ^{**})	174(48.60% ^{**})
4	226(32.75% ^{**})	64(29.63% ^{**})	94(31.65% ^{**})	204(34.46% ^{**})	58(24.27% ^{**})	85(23.74% ^{**})
3	151(21.88% ^{**})	43(19.91% ^{**})	48(16.16% ^{**})	102(17.23% ^{**})	35(14.64% ^{**})	65(18.16% ^{**})
2	41(5.94% ^{**})	8(3.70% ^{**})	11(3.70% ^{**})	18(3.04% ^{**})	12(5.02% ^{**})	19(5.31% ^{**})
1	26(3.77% ^{**})	14(6.48% ^{**})	10(3.37% ^{**})	25(4.22% ^{**})	22(9.21% ^{**})	15(4.19% ^{**})

*: The result is the value for each rating and item divided by total of each item

The results show that there is a trend when the results of the evaluation of each function are compared to the labeled data. It was found that a lower score for each evaluation resulted in a higher percentage of comments being written about the feature being evaluated compared to the other features.

5. CONCLUSION

In this study, we used text mining and machine learning to create a system that uses customer impression of score rating and comments to label a product purchase site. The system is generic in its overall structure. In 3.2.2 and 3.2.5, information specific to Sony TVs is needed. The results extracted using the system were then used to study trends in score ratings and customer impressions. As a result, it was found that the overall rating score was higher than 4.1, because the overall rating is composed of the people who purchased the product in question. When looking at the ratings for each item, the overall rating score was high, and while they were satisfied or generally satisfied, they were not satisfied with some of the features. However, the overall rating score is high, as they are generally satisfied with the products they have purchased.

The overall accuracy of machine learning was 77.90% compared to 4.2, but the accuracy of Picture Quality was very low, with only 36.29% of the results. Many of them had been categorized as predictive of Sound Quality. Therefore,

the results of machine learning were not used to analyze the labeling and evaluation results. In addition, when the results of labeling from 4.2 were compared with the results of the evaluation of each evaluation feature, it was found that if the evaluation results of each evaluation feature were poor, the comments tended to include more information about that evaluation feature. Since the evaluation results for Easy Use and Sound Quality were poor for each feature, the analysis focused on those two features. The results showed that the percentage of respondents who gave a rating of 1 or 2 was higher in Easy Use than in Other. For Sound Quality, the percentage of respondents who gave a rating of 1 or 2 was the highest. Also, given the large percentage of Other, it is possible that they have opinions beyond the Easy Use, Design, Features, Picture Quality, and Sound Quality that was available at this time.

However, this study did not consider each product or model in this study because it was analyzed in terms of Sony TVs. In other words, it doesn't consider whether the target TV is for gaming, watching TV on a TV station, or watching a movie. Since TVs often have different features for different makers and models, the results may vary depending on the maker and model. What can be said from this study is that when comparing the overall rating and feature rating for Sony TVs, the overall rating is high, but Easy Use, Features, and Sound Quality are not rated high, and the other scores are

linked. Also, if the feature rating is poor, the text will often mention that feature rating.

Future efforts include reviewing the labeling step in 3.2.5 to improve the accuracy of the model and addressing the need to improve the accuracy by reviewing the model cre-

ating in 3.2.7. we will also use this study to validate it for other manufacturers and other product. In addition, we try to understand how customers evaluate the product and what they are dissatisfied with the product. Furthermore, we are considering construct a system to show the indicators for product improvement using the extracted information.

REFERENCES

- [1] Ministry of Internal Affairs and Communications in Japan: White Paper in Information and Communications, Internet Usage. 2016.
- [2] ITU: Global numbers of individuals using the Internet total and per 100 inhabitants, 2001-2014. 2014.
- [3] ITU: Global numbers of individuals using the Internet total and per 100 inhabitants, 2005-2019. 2019.
- [4] Ministry of Internal Affairs and Communications (MIC): Survey of Household Spending, Online Shopping. 2017.
- [5] Ministry of Internal Affairs and Communication in Japan: White Paper on Information and Communication: Items Purchased and Reasons for Use in Online Shopping. 2016.
- [6] Ministry of Internal Affairs and Communication in Japan: Study on the Social Impact of the Evaluation of ICT. 2014.
- [7] Hashimoto Y, Ueno S, Ouchi N. Quantitative understanding of product value added using text mining, Japan Society for Management Information Studies Fall Notional Research Presentation Meeting. 2015.
- [8] Jansen BJ, Zhang M, Sobel L, et al. Twitter power: Tweets as electronic word of mouth. *Journal of the American Society for Information Science and Technology*. 2009.
- [9] Tsujii K, Tsuda K. A method of extracting attention information from lodging reviews using text mining. *Digital Practice of Information Processing Society of Japan*. 2012.
- [10] Kim SM, Pantel P, Chklovski T, et al. Automatically assessing Review Helpfulness, the 2006 Conference on Empirical Methods in Natural Language Processing. 2006: 423-430.
- [11] Naik A, Samant L. Correlation review of classification algorithm using data mining tool: WEKA, Rapidminer, Tanagra, Prange and Knime. *International Conference on Computational Modeling and Security*. 2016.
- [12] Alsqer AF, Sasi S. Movie review summarization and sentiment analysis using rapidminer, *International Conference on Networks & Advances in Computational Technologies(NetACT)*. 2017.
- [13] Chauhan S, Grace GH. Document Clustering with Semantic Analysis using RapidMiner. *International Journal of Pure and Applied Mathematics*. 2018; 118(23): 525-533.
- [14] Dhanalakshi V, Bino D, Saravanan AM. Opinion mining from student feedback data using surprised learning algorithms. 3rd MEC International Conference on Big Data and Smart City(ICBDSC). 2016. <https://doi.org/10.1109/ICBDSC.2016.7460390>
- [15] Siregar AM, Faisal S, Tukino T, et al. Comparison Study of Term Weighting Optimally With SVM In Sentiment Analysis. *The 2nd International Conference On Advance And Scientific Innovation(ICASI)*. 2019.
- [16] Moraes R, Valiati JF, Wilson P. Document-level sentiment classification: An empirical comparison between SVM and ANN. *Expert System with Applications*. 2013; 40(2): 621-633. <https://doi.org/10.1016/j.eswa.2012.07.059>
- [17] Pachecp RG, Hamilton D, Myers B. KNN vs. SVM: A Comparison of Algorithms. *Idaho Conference*.
- [18] Harjule P, Gurjar A, Seth H, et al. Text Classification on Twitter Data, *Machine Learning and Internet of Things(ICETCE)*. 2020.