PAPER • **OPEN ACCESS**

Measuring e-Commerce service quality from online customer review using sentiment analysis

To cite this article: Puspita Kencana Sari *et al* 2018 *J. Phys.: Conf. Ser.* **971** 012053

IOP Conf. Series: Journal of Physics: Conf. Series **971** (2018) 012053 doi:10.1088/1742-6596/971/1/012053

Measuring e-Commerce service quality from online customer review using sentiment analysis

Puspita Kencana Sari[1], Andry Alamsyah[2], Sulistyo Wibowo[3]

School of Economics and Business Telkom University, Bandung, Indonesia
puspitakencana@telkomuniversity.ac.id[1], andrya@telkomuniversity.ac.id[2], sulistyo.wib@gmail.com[3]

Abstract. The biggest e-Commerce challenge to understand their market is to chart their level of service quality according to customer perception. The opportunities to collect user perception through online user review is considered faster methodology than conducting direct sampling methodology. To understand the service quality level, sentiment analysis methodology is used to classify the reviews into positive and negative sentiment for five dimensions of electronic service quality (e-Servqual). As case study in this research, we use Tokopedia, one of the biggest e-Commerce service in Indonesia. We obtain the online review comments about Tokopedia service quality during several month observations. The *Naïve Bayes* classification methodology is applied for the reason of its high-level accuracy and support large data processing. The result revealed that personalization and reliability dimension required more attention because have high negative sentiment. Meanwhile, trust and web design dimension have high positive sentiments that means it has very good services. The responsiveness dimension have balance sentiment positive and negative.

1. Introduction

Recently, Internet technology make people can easily generate online contents and able to share information quickly, such as in the review site, e.g: TrustedCompany, Tripadvisor; and in social media e.g: Facebook, Twitter. User generated content sites that review about a product or service from a company can be made as one of the valuable data sources for the company. The activity of data mining from those sites can be an important insight for the company.

Data analytics is an interdisciplinary field with the objective to find pattern on the data with contributions from many fields, such as statistics, machine learning, information search, pattern recognition, and bioinformatics [1]. This activity is widely used in many fields, such as in retail, finance, telecommunications, and media. The several techniques or model for data analytics including classification, prediction, clustering, outlier detection, association rules, sequence analysis, time series analysis, and text mining, as well as new approach such as social network analysis and sentiment analysis [2]. Research in data analytics or specifically data mining, has successfully led to methods, tools and algorithm to solve user's problems using many data type to apply on the following applications, such as business intelligent, predictive model, and decision support system.

To understand their service quality level, most companies use conventional methods to obtain data using sampling and/or questionnaire. This approach has limitation in terms of taking longer time and higher operation cost. We propose a different approach to analyze service quality level using user generated content data from online review site. Because of unstructured nature of the data, we use sentiment analysis method [2] to classify service quality dimensions according to users' perception. By using this method, we can determine the perceived e-commerce service quality.

In this paper, we use Tokopedia, one of the largest e-commerce in Indonesia as our case study to be evaluated its service quality. Between 2016-2017, Tokopedia has 50 million website visitors' and sales number achieve 16.5 million every month [3]. The general purpose of the study is to determine the services quality of e-commerce site using Tokopedia users' review on the TrustedCompany, an online review site. The objective of this research is to know how consumer perception based on data from TrustedCompany using service quality measurement, which is e-service quality. The results of this study can be used as an evaluation for Tokopedia about their service quality.

2. Theoretical Background

Service quality (servqual) is a model to assess services based on its performance's perception. There are five general service quality dimension: *tangibles, reliability, responsiveness, assurance,* and *empathy* [4]. There are some issues in servqual model usage in some industry, one of the issue is in handling service of electronic business or using electronic media [4]. The conceptual model to understand and to improve the service quality in electronics form is called e-Servqual [5]. The new model suggests that e-Servqual incorporate new dimension according to the scale required to measure the quality of e-commerce services:

Table 1. e-Servqual Dimensions [5]

Web Design	Web site design describes the appeal that user interface design presents to customers.
Reliability	Reliability represents the ability of the web site to fulfill orders correctly deliver promptly and keep personal information secure.
Responsiveness	Customers expect online stores to respond to their inquiries promptly
Trust	Trust is defined as customer willingness to accept vulnerability in an online transaction and based on their positive expectations regarding online store behaviors
Personalization	Personalization involves individualized attention personal thank you notes from online stores and the availability of a message area for customer questions or comments

Sentiment analysis is a field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes, and emotions to entities such as products, services, organizations, individuals, problems, events, topics, and their attributes [2]. Generally, there are three levels of sentiment analysis, they are document level, sentence level, entity and aspect level [6]. Sentiment analysis performed automatically to classify opinions into classes by contained words. Multiclass classification is a classification method which has more than two target classes.

Pre-processing is preparation step to process data before entering the classification process [7]. Data resulted from pre-processing step is required to guarantee model fitness and result accuracy. The target class is all e-Servqual dimensions. The details of pre-processing step are as follows.

1) Tokenization process : to divide sentence into words, phrase, or symbol.
2) Filtering process : to remove stopwords or unnecessary text attributes such as *'dan', 'atau', 'dari', 'untuk'*.
3) *Stemming* process : to change words into their word stem, or basic such as '*berjualan'* to *'jual', 'membeli'* to **'***beli'*. We distinguish the stopwords dictionary in *Bahasa* between binary classification, positive-negative sentiment classification case and Multiclass Classification, and five dimension e-Servqual classification case.

Table 2. Pre-processing Example

Raw Data	barang kurang update terbaru dan saat mau pembayaran membingungkan jd gk beli dech
Tokenization	barang, kurang, update, terbaru, dan, saat, mau pembayaran, membingungkan, jd, gk beli, dech,
Filtering Stopword	Barang kurang update terbaru pembayaran membingungkan gk beli
Stemming	Barang kurang update baru bayar bingun tidak beli

After finishing the pre-processing procedure, each word given weight with *TF-IDF* methods. The *TF-IDF* is a method to weighting words which is combination step between the *Term Frequency*, and the *Inverse Document Frequency*. *TF-IDF* give higher weight to words that have low frequency of occurrence in some documents, and simultaneously, has high frequency of occurrence in a document [8]. This ensures that words with high *TF-IDF* values can be used as representative examples of the documents in which they originate, and words that often appear in documents, will be given a low weight.

Naïve Bayes Classifier (NBC) is a statistical classification method that can be used to predict class membership probabilities, such as the probability that a tuple is assigned belonging to a specific class [9]. The NBC's advantages are easy to implement, requires little training data for parameters estimation, and higher accuracy in many studies [10].

3. Methodology

In this research, we use dataset collected comes from an online review website namely Trustedcompany which reviews e-Commerce sites in *Bahasa*. The contents are all reviews regarding *Tokopedia*. The data collection period is from January 2017 until April 2017. As the result, we get total 609 data reviews. The data that has been taken is cleaned through the pre-processing process. Unnecessary attributes are discarded, we only keep data that compatible with research needs. After we get the data, we construct machine learning model to automate the classification process with help of *Rapidminer* software.

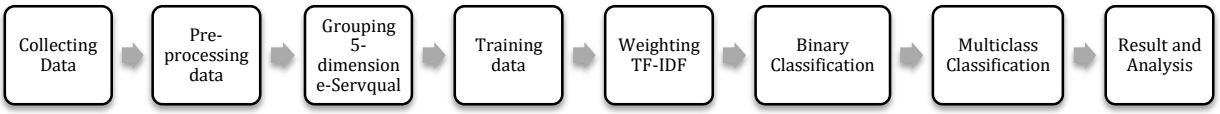

Figure 1. Research Workflow

We classify user review based on five e-Servqual dimensions and sentiment analysis. The five dimensions e-Servqual classification example can be seen in Table 3. Meanwhile, Table 4 is sentiment analysis classification example. To measure model accuracy, we use several measurement, such as: accuracy, recall, precision. We also calculate the confusion matrix [11] shown in Table 5. *TP (True Positive)* is the positive tuples that were correctly labelled by the classifier, *FP (False Positive)* is the negative tuples that were incorrectly labelled as positive, and *TN (True Negative)* is the negative tuples that were correctly labelled by the classifier, and *FN (False Negative)* is the positive tuples that were mislabelled as negative.

Table 3. Example of e-Servqual Dimension

Reviews	*eServqual Dimension*
Sulit berhubungan dengan produsen. utk komunikasipun kadang berbelit2 dan lama.	*Personalization*
Barang sesuai pesanan. Kiriman tidk pernah terlambat	*Reliability*
Pedagang disana dikirimin pesan pada gak responsive.	*Responsiveness*
Pembelian barang terjamin, aman transaksi belanjanya, dan mudah.	*Trust*
webnya sudah sangat baik dan nyaman.	*Web Design*

Table 4. Sentiment Analysis Classification Example

Reviews	*Sentiment*

Hari ini sangat mengecewakan tokopedia. Status sdh di kirim 2 jam yg lalu tp msh blm di terima. Padahal di kirim by gosend.	*Negative*
Bagus banget buat segala kebutuhan one stop solutions	*Positive*

Table 5. Confusion Matrix

	***True* Positive**	***True Negative***
***False* Positive**	42 (TP)	2 (FP)
***False* Negative**	8 (FN)	48 (TN)

The equation of accuracy, recall, precision formula is as follows

$$Accuracy = \frac{TP+TN}{TP+FN+TN+FP} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

From formula (1) we get Accuracy result is 90%, from formula (2) we get Precision is 90,5%, and from formula (3) we get Recall is 90%. Indicating that machine learning model using NBC method successfully classified the data.

4. Result and Analysis

From 609 review data, we classify 321 positive sentiment and 288 negative sentiment, shown in Table 6. For the case of multiclass classification of e-Servqual dimension, the result is shown in Table 7. From table 6 data is classified based on e-Servqual dimensions per sentiment category, the result proportion sentiment on each dimension is shown in Table 8.

There are 138 data in reliability dimension. There are not too significant difference between negative (46%) and positive (54%) sentiment. From five dimensions of e-Servqual, two dimension, (Reliability and Responsiveness) have sentiment value almost equal. Meanwhile, other two-dimensions (Trust and Personalization) have contrary sentiment value. Therefore, the whole sentiment value almost balance. This case shown that consumers perception about service quality from website have balance value although Tokopodia has some weaknesses, but also has some strengthness.

Negative sentiment about the Reliability dimension have topics such as; goods are not met buyer expectations and not sent by the seller. The Personalization dimension has the most negative sentiments (79%) among the other dimensions. Topic that often reviewed by *Tokopedia* users is regarding customer service performance to handle complaints. The Trust dimension has the most positive sentiment (92%) among other dimensions. Topics that are often reviewed regarding about transactions security, as *Tokopedia* implements "*escrow*" service as guarantee. The Web Design dimension also has dominant positive sentiment (66%). It means that *Tokopedia* user interface is good. Most users praise the usability, the look, and the easiness to navigate the website. The Responsiveness dimension has a balanced value, 50% for positive sentiment and 50% for negative sentiment.

Table 6. Binary Classification Result

	Number	*Percentage (%)*
Sentiment Positive	321	53
Sentiment Negative	288	47
Total	609	100

Table 7. Multiclass Classification Result

Dimension	*Number*	*Percentage (%)*

Reliability	138	23
Personalization	139	23
Responsiveness	121	20
Trust	112	18
Web Design	99	16
TOTAL	609	100

Table 8. Sentiment Proportion of each e-Servqual Dimension

Dimension	*Number*	*Sentiment Positive (%)*	*Sentiments Negative(%)*
Reliability	138	46	54
Personalization	139	21	79
Responsiveness	121	50	50
Trust	112	92	8
Web Design	99	66	34
Total	609	321	288

5. Conclusion

Sentiment analysis using Naïve Bayes Classifier method perform well to classify sentiment and issue regarding e-commerce e-Servqual from online review site. Data collection through online source is more appropriated for the research object, since the object also conduct their business through online modus. We considered this approach are more effective and efficient. Moreover, the methods to process unstructured data such as text is available and robust. However, this procedure need to be investigated thoroughly as we collect the data during specific duration. In order to get more representative result we need to conduct another source data such as social media (Twitter and Facebook) to verify the service quality.

The implication from this research is company can evaluate their electronic service quality from consumer perception faster and cheaper because it exploit data which has existed from website review such as Trustedcompany or review in their website. Difference with conventional way, company have to collect data using questioners to evaluate service quality.

Based on the results, the overall sentiment about *Tokopedia* service quality is positive. We conclude *Tokopedia* service quality is good enough. Although the margin is not too high between positive and negative sentiment. On the e-Servqual dimensions, the personalization dimension should get serious attention because they are dominated by negative sentiments.

References

[1] Alamsyah, A., Rahma, W., & Irawan, H. (2015). *Sentiment Analysis Based On Appraisal Theory For Marketing Intelligence In Indonesia's Mobile Phone Market. Journal of Theoretical and Applied Information Technology (2015)*

[2] Zhao, Y. (2013). R Data Minning: Examples and Case Studies. *R and Data Mining,* 2012

[3] Tempo.co. (2016). *Tokopedia Jadi Situs E-Commerce No 1 di Indonesia.* [online]. https://bisnis.tempo.co/read/news/2016/08/17/092796716/tokopedia-jadi-situs-e-commerce-nomor-1-di-indonesia

[4] Zeithaml, V.A. (2002), "Service excellent in electronic channels", Managing Service Quality, Vol. 12 No. 3, pp. 135-8.

[5] Lee, Gwo-Guang., & Lin, Hsiu-Fen. (2005). *Customer Perceptions of e-service quality in online shopping, 33(2). International Journal of Retail& Distribution Management.*

[6] Liu, B. (2012). *Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Language Technologies, 5*(1), 1–167.

[7] Kiz, Aysun Kapucugil., dan Özdağoğlu, Güzin. (2015). Text Mining As A

Supporting Process for VoC Clarification. *Alphanumeric Journal*, 3(1), 025–040. Retrieved from The Journal of Operations Research, Statistics, Econometrics and Management Information Systems

[8] Alamsyah, A., & Saviera, F. (2017). *A Comparison of Indonesia's E-Commerce Sentiment Analysis for Marketing Intelligence Effort (case study of Bukalapak, Tokopedia and Elevenia). The* 8th International Conference on Sustainable Collaboration in Business, Technology, Information and Innovation (2017)

[9] Moens, M. F., Li, J., dan Chua, T.-S. (2014). Mining User Generated Content. Boca Raton: CRC Press.

[10] Han, J., Micheline K. & Jian P. (2012). Data Mining: Concepts and Techniques (3rd Ed.).Waltham: Morgan Kauffman Publishers

[11] Manning, Christopher D., Raghavan, Prabhakar., & Schütze, Hinrich. (2009). *An Introduction to Information Retrieval*. Cambridge University Press