

## **A Core of E-Commerce Customer Experience Based on Conversational Data Using Network Text Methodology**

**Andry Alamsyah<sup>a</sup>, Nurlisa Laksmiani<sup>b</sup>, Lies Anisa Rahimi<sup>c</sup>**

*<sup>a,b,c</sup> School of Economics and Business, Telkom University, Bandung, Indonesia*

*<sup>a</sup> andrya@telkomuniversity.ac.id*

*<sup>b</sup> nurlisalaksmiani@gmail.com*

*<sup>c</sup> cacabocul@gmail.com*

### **ABSTRACT**

Many Indonesian prefer to shop via e-commerce for the reason of its practical nature. E-commerce provides transparent process, including the possibility to voice customer concern and opinion regarding their shopping experiences. Those experiences provide a meaning to understand customer experiences comprehensively. Many Indonesian customers expressed their voice to open social network services such as Twitter and Facebook, where a large proportion of data is in conversational or unstructured text data form. By exploring those media, we explain the e-commerce services level.

A method for finding core topics in large-scale unstructured text data is needed, where the method should be fast but sufficiently accurate. Processing a large-scale data is a non-trivial task, it often needs special skills of peoples and complex computer system. We propose summarizing procedure based on frequently appeared words and their words association to form network text. This method is adapted from Social Network Analysis by model the relationship between words instead of actors. We apply modularity metric to find topic-based customer experiences regarding B2C and C2C e-commerce business models, where each is represented by two biggest e-commerce sites respectively. The result describes customer experience core topics that need to be addressed.

*JEL Classifications:* C8, C88

*Keywords:* consumer behavior; customer experience; e-commerce; network text; social network analysis

## I. INTRODUCTION

Information and communication technology development force the way people and companies doing business to achieve high efficiency and effective work effort. Those each entity adopts the technology in accordance with their goals and needs. The increasing penetration of internet users in Indonesia every year is one of the impacts caused by technological developments in business and government. Following the world trend, the popularity of online shopping in Indonesia is also rising. Online shopping business is a very attractive opportunity aligned with the growth of Indonesia internet users. The wide range of attractive online products and services offerings also become a factor in accelerating the growth of online shopping transactions. In bigger term, online shopping is known as electronic commerce or E-commerce

Last report status by Dan (2014) shown that Indonesian e-commerce market size is constantly increasing each year. Many e-commerce platforms and ecosystem also offer an opportunity for the small-medium enterprise to be part of the economic growth. Some factors such as effectiveness, efficiency, and transparency make online shopping activities preferred by many people. In this research, we investigate Business-to-Consumer (B2C) and Consumer-to-Consumer (C2C) e-commerce business model. Our investigation focusses on customers' opinion about e-commerce service level and comparison among competing for e-commerce business. Although customers can leave their shopping experiences on the website or application, most of them are limited to the individual shopping experience. The general opinion can be mined from popular social network services such as Twitter and Facebook, where both are the biggest social media in Indonesia. For the case study, we pick two most popular B2C and C2C e-commerce sites; they are Lazada and MatahariMall as the B2C type and Tokopedia and BukaLapak as the C2C type.

Customers experience toward an e-commerce site contain their own valuation towards the perceived service. There are good and bad experiences for any type interaction regarding the services, either customer made the purchase or just visiting. Santos et al (2007) point out some scenarios accelerate bad experiences including dissatisfactions about unclear information, delayed update information from the seller, payment problem, delayed delivery, extra charges in the purchase related to delivery, receipt of incomplete orders, and damaged product. Alamsyah and Aulia (2017) show that public social media such as Twitter and Facebook are used by many companies as their end communication channel to customers. The easy nature of communications forms up much feedback to the company in the form of suggestions, complaints, and general forms of dissatisfactions. We rarely see general forms of service compliments in Indonesia market characteristics. Thus, we focus on what topics that dissatisfied customers most discuss about. These dominant topics form a core of customer experience that needs to be followed up by the company or to become reference by future customers. Dissatisfied customers who voice their experience on e-commerce site become a benchmark for the site itself to measure their service level. Based on that phenomenon, the government issued a policy about a new e-commerce industry roadmap, which is written in the 14th volume of the economic policy package. The details of the policies are, the consumer protection to harmonize the regulations concerning electronic certification, accreditation processes, payment mechanism policies, consumer protection and e-commerce industry performer, and legal action resolution schemes.

## II. THEORETICAL BACKGROUND

### A. Customer Experience and E-Commerce

According to Hasan (2013), customer experience is a rational and emotional bonding that occurs because of the response to a particular stimulus by optimizing sense (sensory), feel (emotional), think (cognitive), act (action), and relate (relationship) in the marketing efforts before and after the purchase, exchange of information and emotional attachment. Peppers and Roger (2011) state that customers learn about a company through their experience gained after making regular purchases and other interactions. Thus, in addition to an increase in Customer Loyalty. There are also other benefits obtained which are; customers learned more about their own preferences from each company's experience and feedback and the company could learn more about the strengths and weaknesses of each interaction and feedback from their customer experience.

E-commerce is the process of delivering information, products, services, and payment processes through telephone lines, internet connections and other digital access. There are seven common types of e-commerce transactions: Business-to-Business (B2B), Business-to-Consumer (B2C), Consumer-to-Business (C2B), Consumer-to-Consumer (C2C), Business-to-Government (B2G), Government-to-Business (G2B) and Government-to-Citizen (G2C). We focus on B2C, where the seller is a company and the buyer is an individual (Pratama (2015)) and C2C, where the consumers selling to other consumers with the help of online market maker (Landon and Travers (2014)).

### B. Text Mining

Feldman and Sanger (2007) define Data Mining as a knowledge-intensive process in which a user interacts with a document collection over time by using a suite of analysis tools. In a manner analogous to data mining, Text Mining seeks to extract useful information from data sources through the identification and exploration of interesting patterns. Alamsyah et al (2016) state that one of most practical to summarize large scale conversational data is to find frequently appeared words from data sources, which is part of text mining activities. The result of finding frequently appeared words are a visualize from the contents of a topic and show a distribution of the vocabulary from each topic mentioned. Moreover, text mining is also used to identify words association, where the words in a sentence would be represented as having an association or relationship with other words. In this study, we investigate what kind of customer experience gained by customers based on the content of conversations occurred on social media Twitter and Facebook using those methods.

### C. Social Network Analysis

Scout (2000) define Social Network Analysis (SNA) as a study of the relationship of individuals or other social units, such as an organization, to determine the dependence of the behavior associated with social relationships. In this relationship, described in a node and link. Node is an actor in a network and the link is a line connecting a node with other nodes.

We formulate SNA model as graph  $G(N,E)$  where  $N = \{n_1, n_2, \dots, n_j\}$  is a set of nodes and  $E = \{e_1, e_2, \dots, e_j\}$  is a set of edges.  $|N|$  is a number of nodes in the network and  $|E|$  is a number of edges in a network. The network has some attributes or certain properties that can be calculated and analyzed. The properties of this network are used to determine the model of a network and analyze it with any other network model called network property. We show several network properties formulation used in this measurement in Table 1.

Alamsyah et al (2017) show the usage of *Modularity* metric to measure how distinct groups formed in the network. Larger modularity value means the clearer boundary between groups in the network. Each group represents certain conversation context, in this paper means certain customer experience topic. The modularity formula is shown in the equation below:

$$M = \frac{1}{2m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) \delta(C_i, C_j) \quad (1)$$

where  $M$  is modularity value,  $A_{ij}$  is number of relations inside the group,  $k_i k_j / 2m$  is the expected number of relations between word  $i$  and word  $j$ ,  $\delta(C_i, C_j)$  is the Kronecker delta coefficient which equals 1 if  $i = j$ , and otherwise is 0

### III. METHODOLOGY AND EXPERIMENTAL

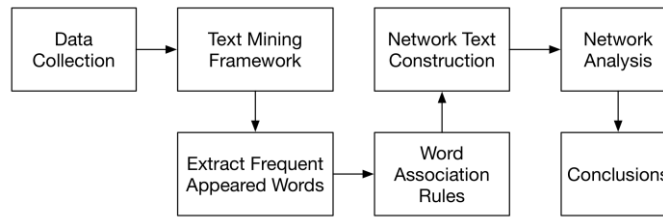
We use text mining to find the patterns between text data, including the relationship between words. Witten *et al.* (2011) explain the special case of data mining into finding the hidden pattern in text relationship, this includes extracting words relationship into meaning for various purposes. We explain the research details in several orderly steps as follows:

1. The data collection step which is data crawling from Twitter and Facebook social network services through their open application programming interface (API) facility. The data are tweets from each e-commerce's Twitter profile and comments from each e-commerce's *Facebook* sites.
2. The text mining framework steps to define the research scope by applying data requirements in preprocessing work. By removal of irrelevant tweets or comments, the objective is to leaves only relevant tweets or comments to the research context, which is about customer experience regarding each E-commerce services in the case study.
3. The frequently appeared words extraction step is to highlight only high-frequency words as a part of pattern recognition effort. We also call the high-frequency words as the dominant words
4. The association rules step is to construct the relationship between the dominant words. The relationship rules defined by the co-occurrence of those words in the same sentence. If two words show up together in a sentence, then they have a relation.
5. The network text construction step accumulates the word association rules into a network of text with the help of SNA model.
6. The network text inherits SNA model characteristics measure network cohesiveness based on its grouping behavior. The grouping quantification metric

called *Modularity* (Newman, 2011). By grouping the text based on their relationship, we can identify the how many topics and how big each dominant topic. The network of topics detection regarding customer experience is essential to improve E-commerce service.

The overall research workflow is shown in Figure 1.

**Figure 1**  
Research workflow



The data profile from each e-commerce service is shown in Table 1.

**Table 1**  
Data profile

	Lazada	MatahariMall	Tokopedia	BukaLapak
Raw Data	25.436 data	3.471 data	10.518 data	10.253 data
Processed	9.100 data	3.514 data	3.692 data	3.210 data
Nodes	200	154	65	65
Edges	619	648	382	305

The association rules show the words pair list shown in Table 2. Due to the limited space, we show only the top 10 of word pair list for each e-commerce service.

**Table 2**  
Top 10 word pairs for each e-commerce sites

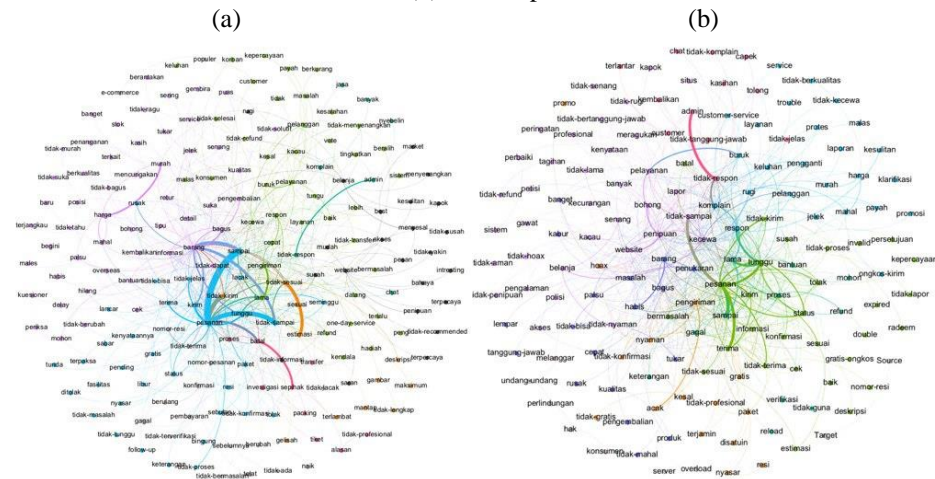
Rank	Lazada Words Pair		MatahariMall Words Pair		Tokopedia Words Pair		BukaLapak Words Pair	
	Words pair	Weight	Words pair	Weight	Words pair	Weight	Words pair	Weight
1	pesanan-sampai	116	pesanan-sampai	56	tidak bisa-	100	dana-belum	88
2	pesanan-tidak	80	pesanan-tidak	49	sampai-sekarang	86	sudah-transfer	71
3	pesanan-tidak	75	pesanan-terima	45	sudah-tunggu	61	sampai-sekarang	65
4	barang-sampai	62	tidak respon-	44	penjual-tidak	58	ada-respon	61

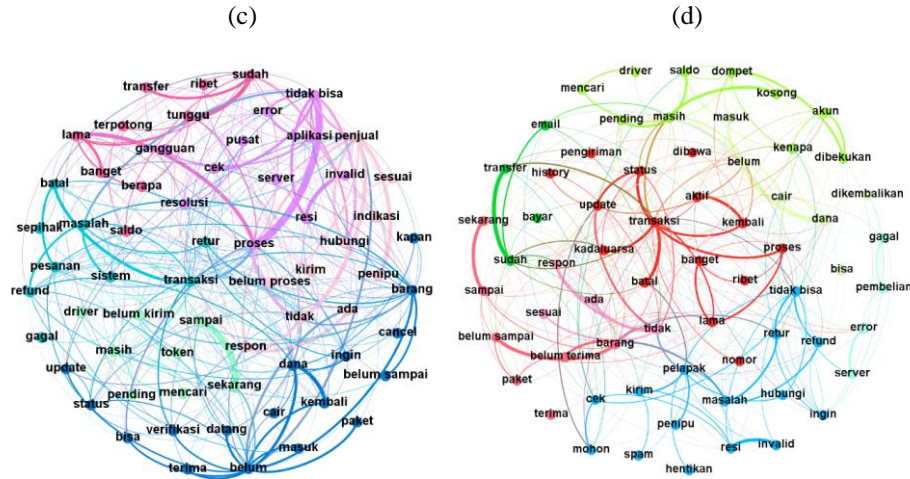
5	tidak sesuai-estimasi	55	tunggu-lama	38	resi-invalid	57	resi-invalid	59
6	barang-tidak sampai	42	tunggu-pesanan	37	dana-kembali	54	transaksi batal	59
7	pesanan-batal	41	tunggu-respon	34	lama-proses	54	barang-belum	57
8	batal-sepihak	36	komplain-tidak respon	25	penjual-penipu	51	masih-kosong	56
9	pengiriman-lama	32	pelayanan-buruk	19	ada-respon	48	transaksi status	51
10	murah-harga	30	pesanan-kirim	19	tidak-proses	45	aktif-kembali	51

IV. RESULT AND ANALYSIS

Figure 2 shows the network text of each e-commerce service. The network has different colors and thickness. Nodes and edges who has the similar color mean they belong to the same group by using Modularity quantification metric. The same group means they belong to the same topic. Edge thickness signifies the weight or how often those word pairs show up together (co-occurrence) in the sentences.

Figure 2  
Network text of customer experience in (a) Lazada, (b) MatahariMall, (c) Tokopedia, and (d) BukaLapak





In Figure 2(a), Lazada has 11 different colors groups which mean there are 11 topics. The top-2 biggest topics shown by green and blue color have their consecutive size is 27.84% and 15.98% of overall network size. The green color refers to “pengiriman” and “lama”, which means the experience felt by customers that the delivery process takes longer than it should be. The blue color refers to “pesanan” and “sampai”, which means that the customer has received their order.

In Figure 2(b), MatahariMall also has 11 different topics. The most dominant topics are shown by the purple and dark colors. Their consecutive size is 20.13% and 12.99%. The purple color refers to “penipuan”, “website”, and “bohong”, which means the customer felt that the MatahariMall website is fraudulently perpetrating them. The dark blue color refers to the word “barang”, “bagus”, and “habis”, which means many good bargain products has already been sold out when the customer made the purchase.

The overall B2C comparison between Lazada and MatahariMall show different issues that regarding customer experiences. The details B2C comparison can be seen in Table 3.

**Table 3**  
Issue comparison on B2C e-commerce sites

No	Lazada	MatahariMall
1.	The two main issues shown based on modularity class are refers to words; “ <b>pengiriman - lama</b> ” and “ <b>pesanan – sampai</b> ”.	The two main issues shown based on modularity class are refers to words; “ <b>penipuan – website - bohong</b> ” and “ <b>barang – bagus - habis</b> ”.
2.	The dominant issue that arises is that <b>the delivery of orders takes a considerable amount of time</b> spent on the Lazada website, but there are some who state that orders received even though they (customers) have to wait for a long time.	The dominant issue that arises is the customer feels that <b>the site is often fraudulent</b> , but they also feel that MatahariMall <b>sells a good quality product</b> .

In Figure 2(c), Tokopedia has 6 different topics. The dominant topics are coming from blue and purple color. They have their consecutive size is 26.56% and 17.19% of overall network size. The blue color refers to “transaksi”, while the purple color refers to “proses” that are often connected. Both colors translated to all customer experience regarding transaction processes, such as the cancelation transaction by the seller, the problem in return and refund process.

In Figure 2(d), BukaLapak, there are 8 different topics. We identify dominant topics shown by red, blue and green colors. Their size is 23.08% for both red and blue colors, and 15.38% for the green color of overall network size. These 3 topics talk about “transaksi”, “status” and “barang”, in which customer experience about issues such as the expiration of transaction status, the transaction status is not quickly updated, the good does not arrive, and the good does not match the description.

The overall C2C comparison shown in Figures 2(c) and 2(d) between Tokopedia and BukaLapak. We can assume that both services have the same issue which associated with their transaction process. The detail of comparison can be seen in Table 4. By learning from customer experience, we have the information that both services need immediate action to resolve their transaction issue. We know that the bad customer experience is likely about service dissatisfaction which is quickly transformed into complaints.

**Table 4**  
Issue comparison on C2C e-commerce sites

No	Tokopedia	BukaLapak
1.	The issues shown based on modularity class are refers to words “ <b>transaksi</b> ” and “ <b>process</b> ”.	The issues shown based on modularity class are refers to words “ <b>transaksi</b> ”, “ <b>status</b> ”, and “ <b>barang</b> ”.
2.	The main issue is directed to the <b>transaction process</b> .	The main issue is directed to the <b>transaction statuses</b> .

From this analysis, we can determine how customer assesses the E-commerce service based on their experience voiced in social media. The proposed topic classification using network text from each E-commerce service simplify huge conversation and review data in social media. The network text also acts as summarization methods; thus, it is easier to detect and to understand large-scale conversational data on social media.

## V. CONCLUSION

Our proposed methodology provides faster result in summarizing the conversational data on social media compared with the previous methodology such as the interview or sampling respondents. It also supplies more detail result compared with online summarization such as word cloud method. Overall this method characteristics accelerate the process to analyze large quantities unstructured data for patterns or knowledge extraction, including customer experience. In addition, network text analysis has the ability to help us to understand the dynamics of the issues, thus measuring it using SNA metrics such as centrality, modularity, path length, and some others.



For general marketing intelligence purpose, this methodology act as complements insight to current methodology. It yields variety results leads to a comprehensive view of customer's voice. At last, the methodology supports e-commerce business to determine accurate future action plan in short time.

## REFERENCES

- Alamsyah, A., E. Sofyan, B.E. Apriliyanti, and V.N. Aini, 2017, "Top Brand Alternative Measurement Based on Consumer Network Activity." *Advance Science Letters*, 23:3813-3816.
- Alamsyah, A., M. Paryasto, F.J. Putra, and R. Himmawan, 2016, "Network Text Analysis to Summarize Online Conversations for Marketing Intelligence Efforts in Telecommunication Industry." *The 4<sup>th</sup> International Conferences on Information and Communication Technology*.
- Alamsyah, A., and A. Aulia, 2017, "Social Network and Sentiment Analysis for Social Customer Relationship Management in Indonesia Banking Sector." *Advance Science Letters*, 23:3808-3812.
- Dan, C, 2014, "Consumer-To-Consumer (C2C) Electronic Commerce: The Recent Picture." *International Journal of Networks and Communications*, II, 29-32.
- Feldman, R., and J. Sanger, 2007, *The Text Mining Handbook; Advanced Approaches in Analyzing Unstructured Data*. Cambridge University. UK: University Press.
- Fliervoet, J.M., G.W. Geerling, E. Mostert, and A.J. Smits, 2015, "Analyzing Collaborative Governance through Social Analysis: A Case Study of River Management along River in The Netherlands." *Environmental Management*, 1-13.
- Hasan, A., 2013, *Edisi Baru Marketing*, Yogyakarta: Media Presido.
- Laudon, K.C., and C.G. Traver, 2014, *E-commerce, Tenth Edition*, United States of America: Pearson Education, Inc.
- Newman, M.E.J., 2011, *Network: An introduction*, University Michigan, and Santa Fe Institute. Oxford University Press
- Peppers, D., and M. Roger, 2011, *Managing Customer Relationships. A Strategic Framework. Second Edition*, Canada: John Wiley & Sons, Inc.
- Pratama, I.P.A.E., 2015, *E-Commerce, E-Business, and Mobile Commerce*, Bandung: Informatika.
- Santos, C.P., and D. Fernandes, 2011, "Failure and Recovery in Internet Purchasing: Their Impact on Consumer Trust and Loyalty to the Company's Site and to the Online Shopping Environment." *Braz. Adm. Rev.* Vol 8 No. 3.
- Scott, J., 2000, *Network Analysis: A Handbook*, Sage Publications.
- Sekretariat Kabinet RI, 2016, "Opening and Developing the Potential of E-commerce in Indonesia." [Online]. <http://www.presidentri.go.id/topik-aktual/membuka-dan-men-gembangkan-potensi-e-commerce-di-indonesia.html>.
- Witten, I., E. Frank, and H. Mark, 2011, *Data Mining: Practical Machine Learning Tools and Techniques*. Third Edition. United States: Elsevier.