Evolving Customer Experience Management in Internet Service Provider Company using Text Analytics

Andry Alamsyah[1], Earlyan Abdiel Bernatapi[2]
School of Economics and Business
Telkom University
Bandung, Indonesia
andrya@telkomuniversity.ac.id[1], earlyanabdiel@gmail.com[2]

***Abstract*— Customer experience is of crucial significance to the constant growth of a business. It is necessary to ensure great customer experience, thus maintaining customer loyalty and satisfaction. An approach that intended to develop and improve customer experience is called Customer Experience Management (CEM). CEM is a strategy practiced to track, supervise, and arrange all synergy to help a business focal point on the needs of its customers. This research uses sentiment analysis and topic modeling to analyze the experience of Internet Service Provider customers. The output of this research expected to drive the strategies change in CEM. This research uses data taken from customer tweets on Twitter. It is considering that the data on social media is enormous and unstructured. Therefore, classification using Naive Bayes Classifier applied to assist and expedite in the sentiment analysis process. The classification for sentiment analysis using NBC gained accuracy above 82%. Hence, the classification models using NBC achieve excellent capability for sentiment analysis. To determining topics that often discussed by customers, this research uses the Latent Dirichlet Allocation models for Topic Modeling.**

Keywords—Sentiment Analysis; Topic Modelling; Big Data; Text Analytics

I. Introduction

Customer experience is a particular and emotional feeback from customers who are directly or indirectly interacting with a company. Direct interaction usually occurs when the purchase initiated by the customer, while indirect interaction often involves advertising, news, unplanned encounters with sales marketing, and word-of-mouth recommendations or criticisms [1]. Customer experience encompasses several aspects of a company’s offering, such as the quality of customer care, advertising, packaging, product and service features, simplicity, and reliability. [2]. The results of the customer experience will be diverse for each individual. Customers perceive satisfaction through experience as responses for services that have been received [3]. Understanding the expectation of the customer is also a crucial aspect of customer experience. Without exception, the customer wants a product or service that works as guaranteed.

Generating an excellent customer experience is essential because it affects customer satisfaction, customer loyalty, influences the customer expectations, and creates emotional bonds with the customers [4]. To be able to provide valuable customer experience; thus, the company created Customer Experience Management (CEM). CEM is a business strategy that focuses on and redefines the business from the customer perspective [5]. Companies are also latching and monitoring on social media like Facebook and Twitter to engage with customers as a part of their Customer Experience Management strategy [6]. According to Wilson [7], some of the best practices of using social media for Customer Experience Management are making emotional connections with customers, create a comprehensive outlook of customer priorities, and leverage customer's information from social media to drive employees enterprise in the continual improvement of customer experience.

Since the data contained in social media is massive and unstructured, is required text analytics, which proficient for extracting information automatically from broad textual data and handle unstructured data [8]. Text analytics regularly applied in industry, academia, web, internet, and other fields [9]. Of the many techniques in text analytics, sentiment analysis and topic modeling are capable of implementing for Customer Experience Management [10] [11].

This research use sentiment analysis for analyzing the customer opinions toward the company's products and services using popular classification algorithms called Naïve Bayes Classifier (NBC). NBC is applied to hasten the sentiment analysis because the data contained in social media is immense. This paper conjointly adds the topic modeling to determine the discussion of popular topics between customers using Latent Dirichlet Allocation (LDA). The topic modeling based on social media, for instance, makes it effortless to understand the responses and discussions among customers in online communities. This research combines both methods where the results of the topic modeling are arranged based on its sentimental. The results are more detailed; thus, it makes obviously for business organizations to identify and discern the topics that usually expressed related to their product and service based on a customer's viewpoint.

The prime intention of this research is imposing text analytics to help the company in developing Customer Experience Management. The company will be able to utilize the results of sentiment analysis and topic modeling, which are derived from the advice, criticism, and complaints to advance the condition of products and services. The research objects are IndiHome and First Media. Both of the companies chosen because they have the largest market share in the Internet Service Provider Industry, with total customers are 5.5 million for IndiHome and 1.1 million for First Media. The data used in this research are tweets from IndiHome and First Media customers that posted via Twitter.

II. Literature Review

A. Customer Experience Management

Customer Experience Management (CEM) is the management of the customer experience, which begin with demands and end with good after-sales service providing real satisfaction through excellent service delivery [12]. Customer Experience Management aims to enhance relationships with customers and build customer loyalty [13]. Practicing Customer Experience Management enables companies to achieve differential advantages, such as create decisive moments of truth for the customer resulting in better customer experience, increase revenues, gain customer retention, and positive customer referrals. [14]. To successfully compete in the industry, a burgeoning number of companies are methodically employing the fundamentals and appliance of customer experience management to strengthen customer loyalty. CEM possibly helps in providing strategic guidance for increasing loyalty and determining precisely which aspect of the customer experience is likely to impact which point of customer profitability [5]. Malviya and Varma [15] define six dimensions that must evaluate by the company to create customer experience improvement in the telecommunications industry, namely:

a. Marketing and sales: good marketing practices and attractive promos carry the possibility to influence consumer decisions to buy products.
b. Brand: the product and service provided by the company ought to fit the expectations of the customer.
c. Billing, charging, and cost-management: provide bill transparency, clarity about the prices pinned on products and services, and provide secure payment.
d. Service quality: network quality is the essential thing in maintaining customer relationships
e. Customer support: powerful customer support capabilities play a vital act in forging customer perception.
f. Product service portfolio: an attractive product and service portfolio firmly impacts to customer decisions.

As Imbug et al. [16] explained in their research that focusing and strengthen the customer experience able to secures customer loyalty, expand the customer base, increase profitability, and ensure that companies survive in the industry. Customer experience appeared as the cornerstone source of competitive benefit and differentiation since each experience is subjective and unique based on the individual and personal encounter [17].

B. Sentiment Analysis

Sentiment analysis notorious as mining opinions or analyzing subjectivity. This method extracts information to analyze its subjectivity or the strength of a word in a text that refers to a particular subject or problem [18]. Sentiment analysis can be formulated for supervised learning by determining into three classes, positive, negative, and neutral. Training and testing data used in the existing research are mostly product reviews. Based on the assumption that the review focused on one product and written by one person. The supervised learning method commonly used to conduct sentiment analysis research is Naïve Bayes Classifier [19].

Sari et al. [20], in their study, use sentiment analysis to determine the level of service quality from online customer reviews by using Naïve Bayes Classifier (NBC) and obtain classification accuracy at 90%. Al-Rubaiee et al. [10] propose a framework that combines a supervised machine learning and sentiment analysis approach to tune the Customer Relationship Management (CRM) via Customer Experience Management (CEM). Supervised machine learning used in their research is NBC. NBC managed to classify the text and generate accuracy at 85%. Alamsyah et al. [21] expressed in their research that Naive Bayes is applicable for sentiment analysis. Naïve Bayes Classifier is based on Bayesian theory, where it determines probabilities found on previous events. It assumed that the occurrence of one with others is independence does not affect one another. The Naive Bayes Classifier has convinced its accomplishment in many courses, regardless of its simplicity and strong assumptions [22]. NBC is suitable to use for numerical or textual data. As it applies to the large data set; then, it does not require complicated iterative parameters for classification [23] [24]. This algorithm allowed the data to be adjusted depending on the needs or personalization [25].

C. Topic Modeling

In simple terms, topic modeling is grouping text data based on a particular topic. Topic modeling performs like clustering, by grouping documents based on similarity. The documents extracted into topics to find out the topics that often discussed. Topic modeling is a suite of algorithms that uncover the hidden thematic structure in document collections. These algorithms help to develop new ways to search, browse, and summarize extensive archives of texts [26]. Topic modeling is important to determine the major topic in customer reviews, make it effortless to recognizing the customer perception [27]. Latent Dirichlet Allocation (LDA) has a significant impact on natural language processing and statistical machine learning and has promptly turn into one of the most acclaimed probabilistic topic modeling techniques in machine learning [28]. LDA as a generic probabilistic model taken from the corpus. Primarily, the document will be presented randomly based on the frequency of topic appearance, in which the characteristics of each topic will be distributed through words [29].

In research conducted by Putri et al. [30], they proposed sentiment analysis with a probabilistic topic model using the LDA method to evolve the reviews from tourists to become the specific topic, which allocated into positive and negative sentiments. Wang [31] found that his model using a combined sentiment LDA and topic LDA is more effective than just topic sentiment analysis. Xiong et al. [32] propose a joint sentiment-topic model for the short text reviews, detecting sentiments and topics simultaneously from the text, especially considering the sparse text problem. Topic modeling used by Calheiros et al. [11] to gather relevant topics that characterize a given hospitality issue by a sentiment. Suggests that the results of such research may help managers to improve customer experience.

The LDA graph model is as follow:

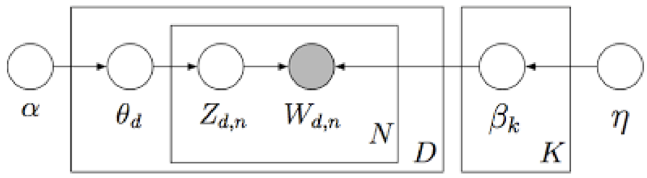

Fig. 1 LDA Graph Model

As Figure 1 illustrated, first, α and η are proportion parameters and topic parameters, respectively. The topics are $\beta 1$*: K*, where each βk is a distribution over the vocabulary. The topic proportion for the *d th* document is θ *d*, where θ d,k is the topic proportion for topic k in document d. The topic assignments for the *d th* document are *Zd*, where *Zd*,n is the topic assignment for the *n th* word in document d. Finally, the observed words for document *d* are *wd*, where wd,n is the *n th* word in document *d*, which is an element from the fixed vocabulary [26].

III. Methodology

The research data fetched from Twitter by using the *Application Programming Interface (*API). The data obtained are 4,122 tweets for IndiHome, while for First Media are 3,127 tweets. There are totals of 7249 tweets collected. Tweets that not included in the criteria for use as training data and testing data are tweets containing ads, not opinions, not included sentiments, spam tweets, and tweets by bots. For IndiHome, totals appropriate data used for the dataset are 2399 tweets, and for First Media are 1804 tweets. The steps process of this research shown in the following figure:

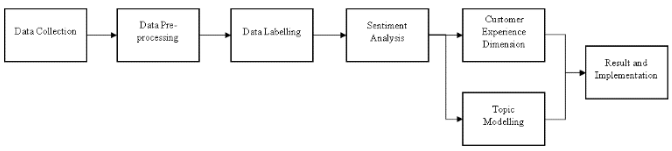

Fig 2. Research Stages

A. Data Pre-processing

In order for the machine to be able to do the classification, it is necessary to do the pre-processing phase. Pre-processing is the initial step to clean up data; therefore, the machine is ready to process data by removing items that are not useful and improving the structure of sentences or words [33]. This stage consists of:

a. Transform Case designed to change upper letters to lower letters.
b. Tokenize is intended to remove posts that are not letters, such as numbers, symbols, and emoji.
c. Filter Token (by Length), this stage serves to filter words depending on the length of the sentence. In this research, the author uses a minimum of 2 words and a maximum of 40 words.
d. Filter Stop-words are using to get rid of words that unnecessary in this research.
e. Stemming: this process aims to convert words in sentences into the root word and eliminate word additions.
f. Generate n-Grams; an n-gram is an arrangement of N in n-gram words [34]. This research is separating each word in the sentence into 2-word sequences.

TABLE I. Implementation of Data Pre-Processing

Pre-processing	Raw Tweet	Result
Case Holding	*@TelkomCare Kenapa IndiHome kalau tanggal tua jaringannya jelek sekali? pdhl sudah melakukan pembayaran bulanan*	*@telkomcare kenapa indihome kalau tanggal tua jaringannya jelek sekali? pdhl sudah melakukan pembayaran bulanan*
Tokenization	*@telkomcare kenapa indihome kalau tanggal tua jaringannya jelek sekali? pdhl sudah melakukan pembayaran bulanan*	*'kenapa' 'indihome' 'kalau' 'tanggal' 'tua' 'jaringannya' 'jelek' 'sekali' 'pdhl' 'sudah' 'melakukan' 'pembayaran' 'bulanan'*
Stopword	*kenapa indihome kalau tanggal tua jaringan nya jelek sekali pdhl sudah melakukan pembayaran bulanan*	*indihome jaringannya jelek sekali pdhl sudah pembayaran*
Stemming	*indihome jaringannya jelek sekali pdhl sudah pembayaran*	*indihome jaringan jelek sekali padahal sudah bayar*
n-Grams	*indihome jaringan jelek sekali padahal sudah bayar*	*'indihome jaringan' 'jaringan jelek, 'jelek sekali' 'sekali padahal' 'padahal sudah' 'sudah bayar'*

B. Data Labelling

After the pre-processing is accomplished, tweets that obtained categorized into two types of sentiments (positive and negative). To understand and ascertain the customer experience in both companies, we categorize the tweet into the dimension, which also proposed by Malviya and Varma [16] in their research, namely marketing sales, brand billing charging, and cost management, service quality, customer support, and product service. This categorizing process is done manually and has the purpose of creating the training data and testing data from each of the Internet service provider's dataset.

TABLE II. Implementation of Data Labelling

Dimension	Positive	Negative
Marketing and Sales	*Saya suka karena banyak promo yang menarik*	*Iklan tidak sesuai dengan apa yang saya dapatkan*
Brand	*Pelayanan yang diberikan sesuai dan tidak mengecewakan*	*Saya berlangganan untuk 50mbps kenyataan hanya mendapatkan 30mbps*
Billing Charging and Cost Management	*Pembayaran bisa dilakukan dengan mudah*	*Tagihan naik secara tiba-tiba*
Service Quality	*Kecepatan internet sangat cepat*	*Jaringan internet dirumah saya sangat lambat*
Customer Support	*Kendala saya dapat ditangani oleh teknisi dengan baik*	*Saya sudah menunggu lama tetapi tidak ada teknisi datang ke rumah*
Product Service Portfolio	*Saya memilih memakai IndiHome karena rekomendasi dari kakak saya*	*Teman saya mengeluhkan layanan dari IndiHome*

C. Sentiment Analysis using Naïve Bayes Classifier

Naive Bayes Classifier employed to automatically classify the sentiment on the dataset. Classification on this research intended to determine positive and negative opinions in tweets. Cross-validation applied in these stages to evaluate the performance of the Naive Bayes Classifier, where the dataset split into two parts, which is 70% for training data and 30% for testing data. Through validation, the process of training will be improved, and its classification results will be more accurate. The training data consist of 2 sentiments, positive and negative.

D. Performance Measurement

For evaluating whether the model works appropriately or otherwise, the performance measurement is required. The performance measurements that conducted in this research are accuracy, kappa, and f-score. Accuracy is used to determine the percentage of positive and negative sentiments that are predicted appropriately. Kappa means calculations to assess the consistency of the results that have been tested. Kappa < 0.4 means low reliability, fair reliability are indicated by kappa at 0.4 to 0.6, while for acceptable reliability are > 0.6 to 0.8, and superb reliability as > 0.8 [35]. F-score is a calculation to figure out the average harmonics of the precision and recall, indicates whether the precision and recall value is good or otherwise. Precision is intended to find out exactly how the classifier prediction compared to the actual data. Recall reveals how complete and sensitive the classifier in predicting data. Precision and recall are considered as perfect when the f score is 1 and will consider as imperfect when it is 0 [36].

E. Topic Modeling using Latent Dirichlet Allocation

We operate Latent Dirichlet Allocation models for topic modeling based on positive and negative sentiment. For flexibly explore the relationships between topic and term, we apply a web-based interactive visualization called LDAvis. The total term frequency is represented by light blue color, while the red color illustrates a predicted term frequency in the elected topic. We determined the topic based on the most extensive-term distribution, the larger the distribution means, the more often the word appears.

IV. Result and Analysis

TABLE III. NAÏVE BAYES CLASSIFIER PERFORMANCE

Measurement	IndiHome	First Media
Accuracy	85.62%	82.76%
Kappa	0.644	0.614
F-Score	90.11%	72.89%

The results obtained from Table III suggest that Naïve Bayes Classifier deserves to be used for sentiment analysis. Naïve Bayes Classifier has the highest accuracy, with 85.62%. This model has good reliability as 0.644 and 0.614. The f-score of Naïve Bayes Classifier is 0.901 for IndiHome and 0.729 for First Media, which means the precision and the recall have the proper rate.

A. Sentiment Analysis Results

TABLE IV. SENTIMENT ANALYSIS RESULT

Internet Service Provider	Positive	Negative
IndiHome	22.00%	78.00%
First Media	24.00%	76.00%

Table IV shows the results of the sentiment analysis on IndiHome, out of 78 percent of customers give a negative sentiment or equals with 1878 tweets, and for the remaining 22 percent of customers respond positively or as many as 521 tweets. While the sentiment analysis results on First Media shown in Table IV above, 76 percent negative sentiment or equivalent to 1369 tweets, and 24 percent positive sentiment or equal to 435 tweets.

TABLE V. CUSTOMER EXPERIENCE OF INDIHOME AND FIRST MEDIA

Dimension	IndiHome (Positive)	First Media (Positive)	IndiHome (Negative)	First Media (Negative)
Billing, charging, and cost management	14	20	261	96
Service quality	345	322	1310	1064
Customer support	124	37	227	159
Marketing and sales	9	30	24	18
Brand	29	26	56	32

According to Table V, for IndiHome, service quality has the most negative response with 1310 tweets, followed by billing, charging, and cost management with 261 responses and the last is customer support with 227 responses. While for First Media, service quality also received the most negative responses with 1064 tweets, following by customer support with responses of 159 tweets, and the last is billing, charging, and cost management with 96 tweets.

In terms of the positive response addressed to IndiHome, the service quality dimension received a response with 345 tweets and subsequently by the customer support with 124 tweets. Whereas for the First Media, service quality still becomes the dimension that received the most positive responses with 322 tweets.

B. Topic Modelling Result

To understand the emerging topics, it is imperative to notice that among the words that presented are interconnected. The issues contained in each topic need to be adjusted to the customer experience dimension that has been proposed to ease the analysis.

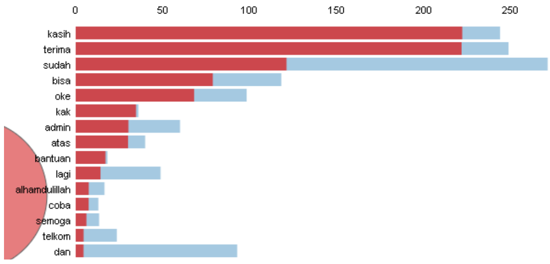

Fig. 3. The positive topic of IndiHome

In Figure 3, the positive topic frequently discussed by the IndiHome customers, concerns about consumer gratitude for support in solving IndiHome services problems. These shown in words *terima, kasih, sudah, bisa, bantuan, Alhamdulillah*. This topic relates to the customer support dimension.

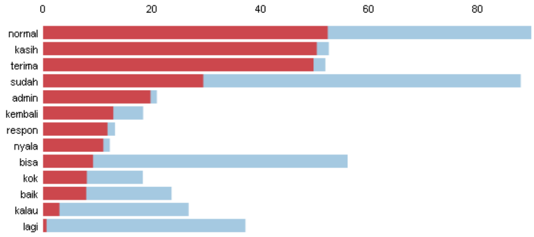

Fig. 4. The positive topic of First Media

In Figure 4, the positive topic for First Media which often discussed regarding the gratitude statement from the customer because the response that has been given to the problem exists, and the service has turned on and returned to normal. These are related to the words normal, *terima, kasih, admin, respon, bisa, dan nyala*. This topic concern about customer support and service quality dimensions.

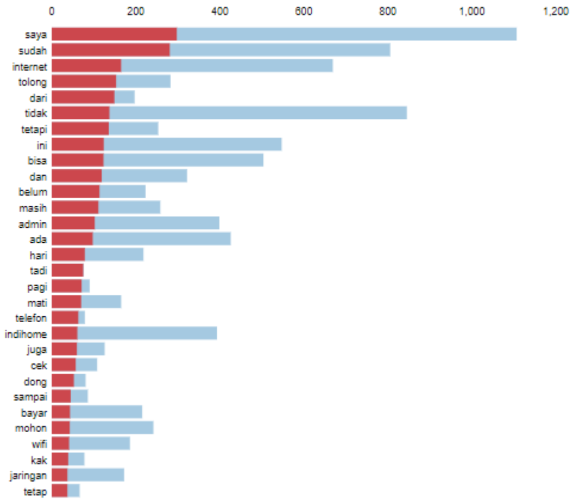

Fig. 5. The negative topic of IndiHome

Figure 5 shows the negative topic about IndiHome. What is discussed in this negative topic is the internet that cannot be used even though payment has been made. The words related to the first topic that can be found in Figure 5 are internet, *belum, mati, bayar*. This topic is associated to billing, charging, and cost management dimensions.

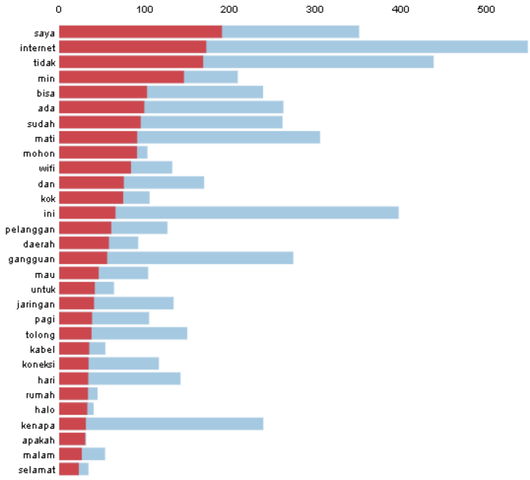

Fig. 6. The negative topic of First Media

Figure 6 shows the negative topic about First Media. The dominant topic, in this case, is about internet networks and television cable that experience interference in certain areas and times. This is related to the words *mati*, internet, *gangguan, rumah, daerah, malam, pagi, jaringan*. This topic pertains to service quality dimensions.

V. Conclusion

By understanding the customer experience, the business organization has a chance to develop their products and services, create marketing innovation, increasing sales, and conducting mutual benefit between customers and companies. This research presents models to discover customer experiences using the combination of sentiment analysis and topic modeling. The intention is to assist and as a principle in conducting Customer Experience Management in the company. It sustains in interpreting the implication of customer experience over time, along with shifts phenomenon.

The sentiment analysis result shows that both companies received more negative sentiment than positive sentiment. These prove that IndiHome and First Media need to improve their customer experience primarily on the quality of service, customer service, and payment system. Topics that often discussed by the customer concerning the variable of customer support, service quality, and billing cost charging management.

Another dimension that related and affects the customer experience could adjoin for forthcoming research. It is advised to add the amount of data to attain precisely classification outcomes. In order to obtain a broad view, additional objects of research from other companies and industries are also proposed. Add another source of data retrieval, such as on websites and other social media besides Twitter. The topic modeling could depend on each customer experience dimension; therefore, results and analysis will provide more detailed.

References

[1] Meyer, Christopher & Schwager, Andre. (2007). Understanding Customer Experience. *Harvard business review.* 85. 116-26; 157.

[2] "Understanding Customer Experience" (2007). *Harvard Business Review*.

[3] Oliver, Riscrd L, (1997), *Satisfaction A Behavioral Perspective on The Consumer*. McGraw-Hill Education, Singapore.

[4] Johnston, R., and Kong, X. (2011), "The customer experience: a road-map for improvement," *Managing Service Quality: An International Journal*, Vol. 21 No. 1, pp. 5-24.

[5] Das, N. & Hota, K. (2014). A Conceptual Study on Customer Experience Management. *International Journal of Management & Business Studies*, Vol. 4, Issue 3.

[6] Parihar, M. (2012). Social Media: The Final Frontier in Customer Experience Management. *Epoch Strategies for Marketing, Family Business, and Entrepreneurship*, *Forthcoming*.

[7] Wilson, D. (2011). Social media customer experience management: The next big trend? My Customer

[8] Moreno, A., & Redondo, T. (2016). Text Analytics: the convergence of Big Data and Artificial Intelligence. *IJIMAI, 3*, 57-64.

[9] Liao, S.H., Chu, P.H., and Hsiao, P.Y. (2012). "Data mining techniques and applications–a decade view from 2000 to 2011," *Expert Systems with Applications*, vol. 39, no. 12, pp. 11 303–11 311.

[10] Al-Rubaiee, H., Al-Omar, K., Qiu, R., & Li, D. (2018). Tuning of Customer Relationship Management (CRM) via Customer Experience Management (CEM) using Sentiment Analysis on Aspects Level.

[11] Calheiros, A. C., Moro, S., & Rita, P. (2017). Sentiment classification of consumer-generated online reviews using topic modeling. *Journal of Hospitality Marketing & Management*, 26(7), 675–693

[12] Seligmann, J. (2018). *Customer Experience Management "THE EXPERIENTIAL JOURNEY."* Lulu.com, United States.

[13] Schmitt B. (2009) Customer Experience Management. In: Bruhn M., Esch FR., Langner T. (eds) Handbuch Kommunikation. Gabler

[14] Joshi, S. (2014). Customer Experience Management: An Exploratory Study on the Parameters Affecting Customer Experience for Cellular Mobile Services of a Telecom Company.

[15] Malviya, R., & Varma, V. (2012). A Value Based Approach to improve Customer Experience, Wipro Council for Industry Research.

[16] Imbug Imbug, N., Ambad, S. N. A., & Bujang, I. (2018). The Influence of Customer Experience on Customer Loyalty in Telecommunication Industry. International Journal of Academic Research in Business and Social Sciences, 8(3), 103–116.

[17] Bagdare, S., & Jain, R. (2013). Measuring retail customer experience. International Journal of Retail & Distribution Management, 41(10), 790–804.

[18] Thelwall, M., Buckley, K., Paltoglou, G., Cai, D., and Kappas, A. (2010). Sentiment strength detection in short informal text. Journal of the American Society for Information Science and Technology. Vol. 61 No. 12, pp. 2544-2558

[19] Bing Liu. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers.

[20] Sari, P. K., Alamsyah, A., & Wibowo, S. (2018). Measuring e-Commerce service quality from online customer review using sentiment analysis. *Journal of Physics: Conference Series, 971, 012053.*

[21] Alamsyah, A., Rizkika, W., Nugroho, D. D. A., Renaldi, F., & Saadah, S. (2018). Dynamic Large Scale Data on Twitter Using Sentiment Analysis and Topic Modeling. *2018 6th International Conference on Information and Communication Technology (ICoICT).*

[22] Roy, D.M. (20016). *Clustered Naïve Bayes*. Department of Electrical Engineering and Computer Science. Massachusetts Institute Of Technology.

[23] Ardhapure, O., Patil, G., Udani, D, and Jetha, K. (2016). Comparative study of classification algorithm for Text based categorization, IJRET: International Journal of Research in Engineering and Technology eISSN: 2319-1163.

[24] Venkata, K.P. (2016). *Emerging Technologies and Applications for Cloud-based Gaming*. IGI Global

[25] Natalius, S. (2011). Metoda Naïve Bayes Classifier dan Penggunaannya pada Klasifikasi Dokumen. Jurnal. Institut. Teknologi Bandung, no. 3.

[26] Blei, D.M., A.Y. Ng, and M.I. (2003). Latent dirichlet allocation. *Journal of machine Learning research.* 2003. 3(Jan): p. 993-1022.

[27] Masrury, R. A., Fannisa, & Alamsyah, A. (2019). Analyzing Tourism Mobile Applications Perceived Quality using Sentiment Analysis and Topic Modeling. *2019 7th International Conference on Information and Communication Technology (ICoICT).*

[28] Tong, Zhou & Zhang, Haiyi. (2016). A Text Mining Research Based on LDA Topic Modelling. Computer Science & Information Technology. 6. 201-210. 10.5121/csit.2016.60616.

[29] Bei Zhao, Yanfei Zhong & Liangpei Zhang. (2013). Scene classification via latent Dirichlet allocation using a hybrid generative/discriminative strategy for high spatial resolution remote sensing imagery. Remote Sensing Letters. 4:12, 1204-1213.

[30] Putri, I., & Kusumaningrum, R. (2017). Latent Dirichlet Allocation (LDA) for Sentiment Analysis Toward Tourism Review in Indonesia. Journal of Physics: Conference Series, 801, 012073.

[31] Wang, W. (2010). Sentiment analysis of online product reviews with Semi-supervised topic sentiment mixture model. in Fuzzy Systems and Knowledge Discovery (FSKD), Seventh International Conference on. 2010. IEEE.

[32] Xiong, S., Wang, K., Ji, D., & Wang, B. (2018). A short text sentiment-topic model for product reviews. Neurocomputing, 297, 94–102.

[33] Thakkar, Harsh & Patel, Dhiren. (2015). Approaches for Sentiment Analysis on Twitter: A State-of-Art study.

[34] Jurafsky, D. & Martin, J.H. (2018). Speech and Language Processing. 3rd ed. Draft. Stanford University.

[35] Landis, J.R., and Koch, G.G. (1977). The Measurement of Observer Agreement for Categorical Data. Biometrics, 33, 159-174.

[36] Shakhovska, Natalia, Medykovskyy, Mykola O. (2018). *Advances in Intelligent Systems and Computing III.* Springer International Publishing. Switzerland.