Review of User Comments for the OVO Fintech application using LDA

Albertus Dwiyoga Widiantoro
Information Systems Department
Computer Science Faculty
Soegijapranata Catholic University
Semarang, Indonesia
email: yoga@unika.ac.id

Bernardinus Harnadi
Information Systems Department
Computer Science Faculty
Soegijapranata Catholic University
Semarang, Indonesia
email: bharnadi@unika.ac.id

Abstract— **The main purpose of this article is to model the topic of OVO Fintech user comments using LDA. This research is a word processing research conducted using text mining. The statistical population is the comments of OVO e-wallet Fintech users in 2021 who take data randomly. LDA (Latent Dirichlet Allocation) and Python programming language were applied to analyze data and implement text mining algorithms from topic modeling. Findings are the most important keywords in Fintech user comments. Also, there are 6 important topics identified in the comments of Fintech users by applying a topic modeling algorithm.Text mining and Latent Dirichlet Allocation were applied to analyze the e-wallet. OVO is a hot topic for Fintech users. Finally, in addition to a retrospective approach to data collection and analysis, the results can be utilized with a prospective approach to strategic planning and policy making.**

Keywords Text mining, Fintech, Topic model, OVO, Latent Dirichlet Allocation, LDA

I. INTRODUCTION

Fintech, also known as financial services/financial technologies/payment technologies [1] is a cross-disciplinary subject that combines finance, technology management and innovation management. Fintech refers to innovative ideas to improve financial service processes by using technological solutions in accordance with existing business models [2]. In terms of technology, Fintech refers to the use of mobile devices including wireless, digital assistants, radio frequency devices, and Near-Field Communication (NFC)-based communication devices to make payments for goods and services [3]. López [4] say that Fintech has higher levels of financial inclusion, not only because of the potential growth-enhancing and poverty-reducing effects of financial inclusion, but also because it contributes to dampening the credit boom-bust cycle if a crisis occurs.

The growth of Fintech in Indonesia is very significant, it can be seen that the Financial Services Authority or OJK (Otoritas Jasa Keuangan) issued information that the number of peer to peer (P2P) based Fintech in 2020 was 155. In research [5] that the growth of financial technology (FinTech) in Indonesia is very significant. From this data, the largest number is Android-based applications installed on the PlayStore.

The Android-based Fintech application in Indonesia has been uploaded to the Playstore at the address https://play.google.com/store. In this study, the comment data for the OVO Fintech application was downloaded using the scraping method.

One of the e-wallet applications in Indonesia is OVO, which is an integrated digital finance application developed by LippoX, integrated with several Lippo companies. As a digital payment company, this application tries to accommodate various needs related to cashless and mobile payments. The OVO app is currently available for Android and IOS platforms. DailySocial's 2021 survey in Indonesia shows OVO to be the most widely used digital wallet application. As many as 58.9% of respondents who use digital wallets admit to using OVO.

Nowadays, smart phones are not only used for communication purposes but also for mobile payment systems (MP), using smart phones allows users to initiate, authorize, and complete financial transactions. MP is very efficient, effective, and convenient for consumers to make transactions because they do not carry cash.

The use of mobile wallets or electronic wallet applications (e-wallet) increases people to make contactless transactions. An e-wallet application using a smartphone is known as a wallet in financial transactions.

The results of the evaluation of comments available on the Internet in the form of positive and negative comments together can affect customer satisfaction. Research on restaurants can use the spontaneous comments of their customers to identify important aspects to manage and improve [6].

Online consumer reviews are very influential in decision making [7], so companies use online consumer reviews to understand consumer demand, the dynamic state of technology, take advantage of commercial opportunities and promote product innovation [8].

In this study, the data used is in the form of text comments from Fintech OVO users in Indonesia. With the research stages: collecting data on OVO users' comments using the Scraping method, Pre-sentence processing, LDA implementation, the topic interpretation on the results of LDA processing on model topics, Conclusions from grouping OVO user comments.

II. Literature Review

Topic modeling, one of the text mining applications, analyzes text to identify themes. This topic model is used to classify documents and to support additional algorithms to perform contextual adaptive feature, fact and relationship extraction [9].

The LDA topic model is a probability generative model that aims to identify key themes in a collection of textual documents [10]. LDA can substantially minimize the number of dimensions (i.e. words) in the text while maintaining the important ties between all dimensions and their main topics in the documents that follow [11]. Based on the occurrence of words, the LDA algorithm simultaneously estimates the topic of each document and allocates it to weights [12]. The result of the topic modeling algorithm is a weighted word list, where each list is a topic and where the higher weighted words in the list better indicate that topic and each document is used as a distribution over topics to detect semantic patterns through all documents.

LDA topic modeling is able to identify research trends and related topics in cluster form. The cluster results have interpreted the emergence of a number of topics in separate clusters and the fact that researchers in the same geographic area cite each other [13].

Topic modeling is one of the most powerful text mining techniques for exploring textual data, detecting latent patterns in text, and finding relationships between textual documents finding hidden topics in document sets [14]. The researchers believe that topic modeling is a useful tool for extracting information from textual data on the one hand and shows better performance in information retrieval compared to many conventional approaches.

In brief, topic modeling is a reliable and practical tool for the study of text mining, and has been discussed in response to the need of researchers to explore in extensive collections of scientific texts and also to introduce a structured and automated approach to identifying the topics present in the body of the text.

The results of topic modeling can analyze how topics are connected and their evolution over time, find topics that are meaningful and related to terms, track them over time and provide an opportunity for the analyst to better understand their relationships and changes.

Topic modeling is one of the most practical and important methods in text mining. Topic modeling can reveal latent integrated frameworks in textual data, and this method helps researchers and planners in macro-decision making [15].

In this section, with LDA topic selection, the topics from the cluster are extracted. Text mining with topic detection is used to extract topics. As a tool to accomplish this task using python.

III. Metodology

The stages of this research are data collection using the Scraping method which leads to the OVO playstore address. Pre-processing sentences by deleting symbols, eliminating unimportant words.

Ones of the unattended Machine Learning algorithms is LDA which identifies latent topics from a collection of documents. This process relies on a "bag of word" approach, which treats each document as a word count vector. Each document is represented as a probability distribution over several topics, where each topic is represented as a probability distribution over a number of words. the stages for each document are processed as follows: For each document, select a topic from its distribution above the topic; one word from the distribution of words related to the topic to be selected; This process is repeated for all words in the user comment document

Weighting, compiling a table according to the weight of each word, interpreting a collection of words into a particular theme or topic. create graphs using pyLDAvis to make it easier to read groups of words into specific topics, and make it easier to see that each group of words does not coincide with each other.

Conclusion of text processing using LDA algorithm

IV. Result and Discussion

The data used is OVO comment data taken from Google Playstore which was taken randomly in 2021. The data structure consists of reviewId, username, userImage, content, score, thumbsUpCount, reviewCreatedVersion, at, replyContent, repliedAt, sortOrder, appId as shown in the table. In this study, we will explore topics from the content column, as shown in table 1.

Sentences that will be observed to be used as topics as shown in the content column. as shown in table 2.

The total data used after cleaning is 1305. The data consists of user comments with a value score of 1 to 5 with a distribution as shown in the table III below. in addition to reviews of the OVO application, users also give a rating of 1-5, where a rating of 1 is considered to have the lowest/hurtest performance, rating 3 means neutral or mediocre, and 5 is considered to have the best performance.

TABLE I. STRUCTURE COMMENT

reviewId	userName	userImage	content	score	thumbsUpCount	reviewCreatedVersion	at	replyContent	repliedAt	sortOrder	appId
gp:AOqpTC	ABP GDT	https://play-lh	Alhamduli	5	36	3.22.1	########	Hai Kak, mohc	##########	most_relevant	ovo.id
gp:AOqpTC	M Fitra Akb	https://play-lh	Good Top	5	0	3.22.1	########			most_relevant	ovo.id
gp:AOqpTC	dh4n Adver	https://play-lh	==> Tidak	5	60	3.22.1	########	Hai Kak, mohc	##########	most_relevant	ovo.id
gp:AOqpTC	Sania Dhere	https://play-lh	Oke baikla	5	1	3.22.1	########	Hai Kak Sania,	##########	most_relevant	ovo.id
gp:AOqpTC	Irabiah	https://play-lh	Aplikasi in	5	2	3.22.1	########			most_relevant	ovo.id
gp:AOqpTC	Arra T-rex	https://play-lh	Selalu pua	5	1	3.22.1	########			most_relevant	ovo.id
gp:AOqpTC	SHM_ RooN	https://play-lh	Alhamduli	5	0	3.22.1	########			most_relevant	ovo.id
gp:AOqpTC	Darmawan	https://play-lh	Kenapa OV	5	37	3.22.1	########	Hai Kak, mohc	##########	most_relevant	ovo.id
gp:AOqpTC	Sri Novianti	https://play-lh	Saya suda	5	1	3.22.1	########			most_relevant	ovo.id
gp:AOqpTC	Yuliadi	https://play-lh	TERIMAKA	5	0	3.22.1	########			most_relevant	ovo.id
gp:AOqpTC	Didit Yakin	https://play-lh	Sangat me	5	0	3.19.0	########			most_relevant	ovo.id
gp:AOqpTC	Enggal Med	https://play-lh	Aku kasih	5	31	3.22.1	########			most_relevant	ovo.id
gp:AOqpTC	Riki Priman	https://play-lh	Alhamduli	5	3	3.22.1	########			most_relevant	ovo.id
gp:AOqpTC	Eka Purnam	https://play-lh	Aplikasi ov	5	8	3.22.1	########			most_relevant	ovo.id
gp:AOqpTC	rommy wija	https://play-lh	Aplikasi OV	5	0	3.22.1	########			most_relevant	ovo.id
gp:AOqpTC	Bonnie And	https://play-lh	masalah s	5	0	3.22.1	########	Hai Kak, Mohc	##########	most_relevant	ovo.id
gp:AOqpTC	ANDIK YUTA	https://play-lh	Setelah di	5	1	3.22.1	########	Hai Kak, terim	##########	most_relevant	ovo.id
gp:AOqpTC	Jovan Prata	https://play-lh	Tolong do	5	10	3.22.1	########			most_relevant	ovo.id
gp:AOqpTC	Tiara Aprilia	https://play-lh	Aplikasi in	5	0	3.22.1	########			most_relevant	ovo.id

TABLE II. CONTENT COMMENT

No.	content	score	appId
1	Alhamdulillah sudah diperbaiki oleh tim ovo terimakasih atas pelayanan cepatnya, Mohon maaf apabila saya menginginkan penanganan yang cepat karena kondisi sedang urgent. ðŸ™	5	ovo.id
	Sudah 3 hari ini saya mau isi saldo ovo,susah banget,di mini market sudah tak bisa isi saldo,mau isi di grab driver harus pesan dahulu jadi penumpangnya,paling banyak bisa dari ATM,tapi kalau punya ATM ngapain pake ovo..intinya...susah isi saldonya	2	ovo.id
	Udah 1 minggu tidak bisa login,sudah di ulang sampe saat ini ,sudah ke cs pun tetap tidak bisa	4	ovo.id
	Aku penguna app ini.tapi sayang topup saldonya sudah Gak bisa di Alfa or indomart. Kenapa begitu??	4	ovo.id
	Merubah Nama Profil Saja Susahnya minta ampun. Sementara ganti fhoto bisa. Telefon cs berkali kali jawabannya itu itu aja. Buang buang pulsa dan waktu. Padahal nama akun saya hanya salah 1 huruf. Masa ga bisa di betulkan dari sistemnya.	2	ovo.id
2	Good Top One OVO! Tidak hanya untuk belanja apapun yg dipermudah segalanya tapi bahkan kita untuk membuka bisnis maupun dari bisnis kecil & bisnis besar disediakan Oleh Dompet digital (OVO), Bahkan buat kalian yg suka investasi di (OVO) juga ada loh, buat Semua Transaksi/Nabung dll Akan di permudah oleh Dompet digital OVO. Tunggu apa lg buruan Download/instal aplikasi OVO!!!	5	ovo.id
3	==> Tidak bisa login. <== => Kirim helpdesk ke ovo.id "sudah" - permasalahan belum terselesaikan sampai detik ini (permasalahan sebelumnya). -Revisi- => Solusinya, install ulang smartphone"reset factory", baru bisa login "permasalahan Terselesaikan" Case	5	ovo.id
4	Oke baiklah, penanganan cepet, skrg ovo premium sy bisa didaftrn ke nmr baru. Otw ganti bintang ðŸ™ Saya udh prnh upgrade ovo dinomor lama, nah trus hp saya ilang bserta isi2nya, akhirnya buat akun ovo baru, nah mau upgrade gabisa mulu karna udah nyangkut dinomor lama, gimana dong admin, tolong solusinya hubungi saya ya..	5	ovo.id

TABLE III. NUMBER OF COMMENTS BASED ON COMMENT WEIGHT

score	number
1	206
2	239
3	401
4	40
5	419
total	**1305**

LDA used using genism. The installation process by importing genism and its corpora.

Creating the object for LDA model using gensim library. this process is to create a model, by doing several experiments with until it reaches the non-overlapping topics displayed using LDAvis. And obtained a stable model with a topic of 6 with a number of words of 15.

Setting the model lda_model = Lda(doc_term_matrix, num_topics=total_topics, id2word = dictionary, random_state=100, passes=10, alpha=0.01, eta=0.10, per_word_topics=True, chunksize=100)

The process after building the model is to identify the topics studied by the model. Topics that fit into recognizable categories. To do this, we must first remember what 'topic' actually is in the context of LDA.

The probability distribution of words in the vocabulary is that each topic gives a certain probability for each unique word that appears in the data. Different topics will give different probabilities for the same word:

Looking at the highest probability words in each topic will give an idea of the main theme. Each topic can be identified with several clear themes and all topics are relatively different.

In this study, 6 topics will be set to explore the theme of OVO Fintech users' comments.

In-topics have mixed membership, which means that each document can partially belong to several different topics. For each document, topic membership is expressed as a weight vector that adds up to one; the magnitude of each weight indicates the extent to which the document represents a particular topic.

We will explore in our fit model by looking at the topic distribution for user comments from the data set. In this process it should be found that these comments have the highest weight on the topic whose theme is most relevant to the subject of the review.

Here we will predict the distribution of OVO Fintech topics shown in table IV.

Numbers 1 to 6 in the row indicate the topic, numbers 1 to 9 in the column indicate the word that appears in the topic. Each topic has 9 words that appear with the highest weight. Each topic contains words that have a certain weight. The weights are obtained from the probability process carried out by LDA.

TABLE IV. WORD PROCESSING RESULTS BASED ON WEIGHT AND INTERPRETATION

No	Topic interpretation	1	2	3	4	5	6	7	8	9
1	already logged in using an account but it's long-time(difficulty verifying)	0.046* be able	0.042* already	0.027* login	0.020* account	0.017* please	0.016* even	0.014* long-time	0.014* login	0.013* balance
2	complete conditions but failed transaction	0.050* be-able	0.041* transfer	0.029* already	0.021* bank	0.017* account	0.017* failed	0.015* open	0.014* photo	0.013* ID card
3	upgrade to premiere is problematic	0.058* upgrade	0.031* long	0.025* already	0.023* premiere	0.021* register	0.020* waiting	0.018* can	0.015* account	0.014* must
4	transaction convenience	0.039* transaction	0.036* very	0.030* easy	0.021* pay	0.020* sell	0.016* too	0.015* more	0.015* buy	0.015* piece
5	have done top-up for pre-employment response with customer service (CS) support	0.023* already	0.017* word	0.016* code	0.015* me	0.015* cs	0.015* security	0.014* continue	0.013* pre-employment	0.013* now
6	easy top-up service helps in work	0.029* process	0.020* use	0.020* top-up	0.018* already	0.015* fast	0.015* service	0.015* please	0.014* soon	0.014* work

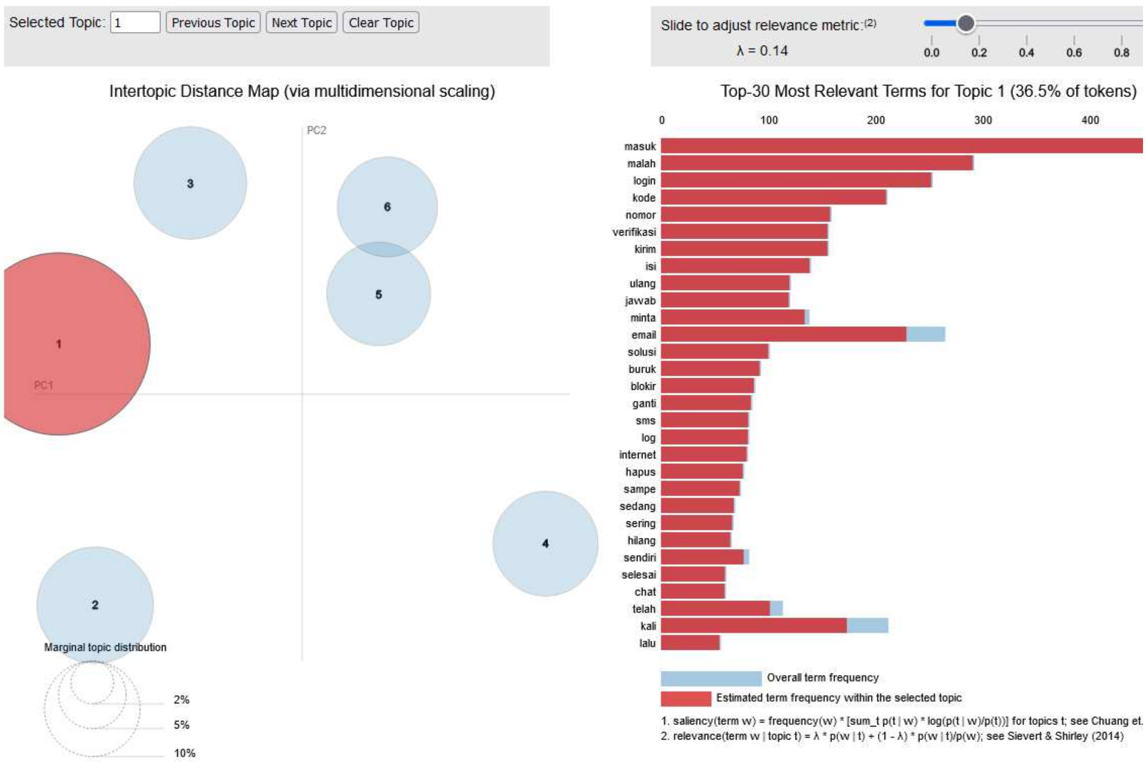

Figure 1. Ovo Comment Data Visualization Using Pyldavis

From table IV, it takes 9 words with the highest weight from each topic that is built. Of these 6 topics, it can be interpreted as a group of topics: difficulty of verification, complete requirements but failed transactions, troubled upgrading to premiere, ease of transactions, doing top-up for pre-employment with Customer Service (CS) response, ease of top-up service helping in work.

By getting an interpretation of the topic, problems or advantages of the OVO Financial Technology application can be identified.

By using LDAvis, this process will visualize using circles that represent certain topics in Fintech.

The gray image shows the overall term frequency; the red image shows the approximate term frequency in the selected topic.

By setting the Lamda, you can see the order of words in the selected topic according to the frequency that is formed.

V. Conclusion

The LDA algorithm can identify hidden topics contained in the comments of Fintech OVO users. The processed data get a maximum topic group of 6 topics with groups that do not appear to overlap which can be shown in the LDAVIS image.

Of the 6 topics, there are 3already logged in using an account but it's long-time (difficulty verifying), complete conditions but failed transaction, upgrade to premiere is problematic. Meanwhile, 3 transaction convenience, have done top-up for pre-employment response with customer service (CS) support, easy top-up service helps in work. From these topics, service providers can focus on 2 parts of service improvement and system repair, on the other hand maintaining and making better services.

Acknowledgment

The highest gratitude goes to the Directorate General of Higher Education, Ministry of Education and Culture of the Republic of Indonesia for providing this research project and thanks to the Information Systems Department of Soegijapranata Catholic University, Indonesia.

References

[1] E. Z. Milian, M. D. M. Spinola, and M. M. De Carvalho, “Fintechs : A Literature Review and Research Agenda,” Electron. Commer. Res. Appl., p. 100833, 2019, doi: 10.1016/j.elerap.2019.100833.

[2] K. Leong and A. Sung, “FinTech (Financial Technology): What is It and How to Use Technologies to Create Business Value in Fintech Way ?,” vol. 9, no. 2, 2018, doi: 10.18178/ijimt.2018.9.2.791.

[3] W. A. Alkhowaiter, “Digital payment and banking adoption research in Gulf countries: A systematic literature review,” Int. J. Inf. Manage., vol. 53, no. February, p. 102102, 2020, doi: 10.1016/j.ijinfomgt.2020.102102.

[4] [4] T. López and A. Winkler, “Does financial inclusion mitigate credit boom-bust cycles?,” J. Financ. Stab., vol. 43, pp. 116–129, 2019, doi: 10.1016/j.jfs.2019.06.001.

[5] D. Hoang, B. Phan, P. Kumar, R. E. Rahman, and A. R. Hutabarat, “Do financial technology firms influence bank performance ?,” Pacific-Basin Financ. J., vol. 62, no. November 2018, p. 101210, 2020, doi: 10.1016/j.pacfin.2019.101210.

[6] G. Tontini, L. Irgang, A. Kroenke, I. Hadlich, J. D. Picolo, and J. Mikulic, “How to use spontaneous customer comments to identify nonlinear background of satisfaction with restaurant services,” Benchmarking, vol. 29, no. 2, pp. 496–521, 2022, doi: 10.1108/BIJ-08-2020-0409.

[7] Y. Li and L. Zhang, “Do online reviews truly matter ? A study of the characteristics of consumers involved in different online review scenarios,” Behav. Inf. Technol., vol. 0, no. 0, pp. 1–19, 2020, doi: 10.1080/0144929X.2020.1759691.

[8] W. Chen, B. Gu, Q. Ye, and K. Xiaoguo, “Measuring and Managing the Externality of Managerial Responses to Online Customer Reviews,” vol. 7047, pp. 1–16, 2018.

[9] J. Al Qundus, S. Peikert, and A. Paschke, “AI supported Topic Modeling using KNIME-Workflows,” CEUR Workshop Proc., vol. 2535, pp. 1–7, 2020.

[10] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet Allocation,” vol. 3, pp. 993–1022, 2003.

[11] M. Ezzeldin and W. El-Dakhakhni, “Metaresearching Structural Engineering Using Text Mining: Trend Identifications and Knowledge Gap Discoveries,” J. Struct. Eng., vol. 146, no. 5, 2020, doi: 10.1061/(asce)st.1943-541x.0002523.

[12] J. L. Hung, “Trends of e-learning research from 2000 to 2008: Use of text mining and bibliometrics,” Br. J. Educ. Technol., vol. 43, no. 1, pp. 5–16, 2012, doi: 10.1111/j.1467-8535.2010.01144.x.

[13] E. Ç. Bilge and H. Yaman, “Research trends analysis using text mining in construction management: 2000–2020,” Eng. Constr. Archit. Manag., 2021, doi: 10.1108/ECAM-02-2021-0107.

[14] H. Jelodar et al., “Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey,” Multimed. Tools Appl., vol. 78, no. 11, pp. 15169–15211, 2019, doi: 10.1007/s11042-018-6894-4.

[15] X. Cheng et al., “Topic modelling of ecology, environment and poverty nexus: An integrated framework,” Agric. Ecosyst. Environ., vol. 267, no. March, pp. 1–14, 2018, doi: 10.1016/j.agee.2018.07.022.