Check for updates

Analyzing Technostress Factors: Aspect-Based Sentiment Analysis for Identifying Causes in Fintech Users Using the Decision Tree Algorithm

Sahra Bilqis Fauziyyah [1], Muhardi Saputra [2], and
Riska Yanu Fa'rifah [3]

[1,2,3] Telkom University, Jl. Telekomunikasi, 40257, Indonesia
[1] sahrabifa@student.telkomuniversity.ac.id

Abstract. Information technology innovation, particularly in Financial Technology (fintech), plays a central role in various aspects of life. Among the fintech services, e-wallets are highly popular in Indonesia. In 2021, OVO was a leading e-wallet; however, in 2022, it experienced a decline, suspected to be caused by technostress. People who experience technostress have negative attitudes and feelings towards technology. This research employs Aspect-Based Sentiment Analysis, using LDA topic modeling to identify four aspects: features, access, service, and security. OVO user reviews from Google Play Store were scraped for data analysis. Sentiment classification using C4.5 Decision Tree with a 75:25 data sharing ratio achieved high accuracies: features (96.79%), access (94.95%), service (92.19%), and security (96.36%). The results aid fintech companies, especially OVO, in addressing user technostress and enhancing user experience and engagement.

Keywords: Fintech, E-Wallet, Technostress, Aspect-Based Sentiment Analysis, LDA, Decision Tree C4.5.

1 Introduction

The innovation of information technology has had a significant impact on human life and has transformed various aspects of life, including the financial services industry. One form of information technology innovation in the financial industry is Financial Technology (Fintech) [1]. Fintech is the fusion of information technology and finance that brings innovative financial services, such as digital wallets or e-wallets [2]. In Indonesia, the growth of fintech users, especially digital wallets, has seen rapid growth in recent years [3].

However, this growth has not always been smooth. One example of a digital wallet application is OVO, which was initially very popular and had many users [4]. However, over time, OVO experienced a decline in the number of users and has been surpassed by several other digital wallet applications [5]. This decline can be attributed to various factors, one of which is technostress [6]. Technostress is the psychological pressure that arises as a result of using new technology or changes in technology usage [7]. In the context of fintech, technostress can arise due to user difficulties in adapting to changes

© The Author(s) 2023
M. D. Sulistiyo and R. A. Nugraha (eds.), *Proceedings of the International Conference on Enterprise and Industrial Systems (ICOEINS 2023)*, Advances in Economics, Business and Management Research 270,
https://doi.org/10.2991/978-94-6463-340-5_9

in application features, security and privacy risks, as well as customer support issues [8].

To understand the impact of technostress on fintech users, this research uses an aspect-based sentiment analysis method on the OVO digital wallet application. The aim of this research is to uncover specific aspects of using OVO that cause technostress among users. This is because ABSA can identify the aspects that contribute most to the technostress felt by users [9]. To obtain relevant aspects, the use of topic modeling is one of the most widely used ways [10]. This research implements topic modeling techniques using the Latent Dirichlet Allocation (LDA) algorithm to identify the most frequently discussed aspects by users on the Google Play Store.

The aspect-based sentiment analysis method utilizes user reviews on the Google Play Store as the data source. These reviews are processed using machine learning approaches to classify them as positive, negative, or neutral. From the sentiment analysis results, this research identifies the aspects that contribute the most to the technostress experienced by OVO users. Furthermore, the sentiment analysis results for each aspect are represented using word clouds to visualize user sentiments. The Decision Tree classification algorithm is also used to predict technostress based on negative sentiments expressed by users towards the aspects identified in the LDA analysis. This research is expected to provide deeper insights into specific aspects of using the OVO digital wallet that contribute to the technostress experienced by users. This information can help fintech companies, such as OVO, to identify and address technostress issues faced by users, thereby enhancing user experiences, and retaining a larger user base.

2 Literature Review

2.1 Financial Technology

Fintech generally aims to attract customers with products and services that are more efficient, transparent, user-friendly, and automated [11]. Meanwhile, according to Bank Indonesia Regulation Number 19/12 / PBI / 2017 defines fintech as the use of technology in the financial system that can produce new products, services, technologies, and/or business models that can affect monetary stability, financial system stability, and/or improve the smoothness, security, efficiency, and reliability of the payment system [12].

A digital wallet also called an e-wallet is one of the fintech products used as a digital payment system that allows users to make electronic transactions or online transactions. In addition, digital wallets can also be used to replace cash in shopping, pay bills such as electricity or water bills, and transfer money [13]

2.2 Technostress

In 1984, Dr. Craig Brod, a clinical psychologist, coined the term technostress which explains that technostress is a contemporary condition resulting from the inability to cope with new computer technology in a healthy way, which can be considered a modern adaptation disease. This condition has two distinct but interrelated modes of emergence: difficulty accepting computer technology and a form of over-dependence on computer technology [7]. In addition, it can also be defined that technostress is a feeling of discomfort that can be caused by technological change or a condition of dependence on technology.

3 Methodology

3.1 Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) introduced by Blei, Ng, and Jordan in 2003 [14] has been widely used in the field of Natural Language Processing (NLP) and applied in various tasks, including word sense and multi-document summarization. LDA is a generative probabilistic version of a corpus. The primary idea at the back of LDA is that documents are represented as a random combination of hidden topics, wherein each topic is characterized with the aid of a distribution of phrases. LDA-primarily based topic modeling strategies had been carried out in herbal language processing, textual content mining, social media analysis, and fact retrieval [15].

3.2 Aspect-Based Sentiment Analysis

ABSA or Aspect-Based Sentiment Analysis is one of the sentiment analysis approaches that focuses on specific aspects or attributes in text to extract sentiment. Unlike sentiment analysis which is used to represent the entity as a whole, aspect-based sentiment analysis covers both the entity and its related aspects [9]. In addition, ABSA is also known as topic-based sentiment analysis, aiming to extract sentiment from specific attributes with respect to specific topics in a document, also commonly referred to as feature-based opinion mining [16]. There are several things that must be considered in performing aspect-based sentiment analysis, including:

1. Aspect Extraction: Aspect extraction aims to identify and retrieve aspects that have been assessed or evaluated.
2. Aspect sentiment classification: Aspect sentiment classification aims to determine whether opinions about the aspects have a positive, negative or neutral sentiment.

3.3 Decision Tree

Decision tree method is one of the widely used methods in data mining to create classification systems based on diverse factors or to develop prediction algorithms for specific outcome variables [17]. The C4.5 algorithm proposed in 1993 by Ross Quinlan is a development of the ID3 algorithm [18]. The development of the C4.5 algorithm involves several improvements, such as using the information gain ratio as a criterion in the selection of separating attributes, not limited to discrete attributes but can also handle continuous attributes, handling incomplete training data with missing values, and performing a pruning process when building a tree to prevent overfitting [19].

3.4 SentiStrength

Sentistrength is a classification technique that assesses the strength of sentiment in short texts using a method based on the lexicon, which is a collection of words with a certain sentiment strength or value [21]. Sentistrength uses a combination of linguistic rules and machine learning algorithms to analyze text and provide a sentiment score. The score is generally a number between -5 to +5, with negative scores indicating negative sentiment and positive scores indicating sentiment [22].

3.5 Research workflow

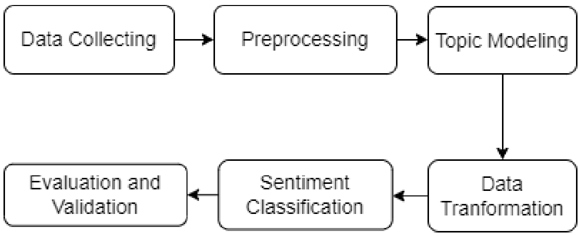

Fig. 1. Research workflow

4 Result and Discussion

4.1 Data Collection

This research collected 100,000 OVO user reviews from the Google Play Store using Python's google-play-scraper library and web scraping technique. The data covered the period from 2018 to 2023, but the researcher narrowed it down to 33,120 reviews from January 1, 2023, to March 14, 2023.

4.2 Preprocessing

The text preprocessing begins with sentence segmentation, separating sentences in reviews based on symbols like periods and conjunction words [23]. This is due to varied sentiments and aspects within each review, making it is not feasible to assign the same label to the whole review. The initial 33,120 reviews increased to 47,467 after preprocessing. Steps include case folding, spelling correction, punctuation and number removal, stemming, stopword removal, and null value removal.

4.3 Topic Modelling

This research uses the LDA model from gensim.models.LdaMulticore to form relevant topics for documents. Text data is transformed into Bag-of-Words and trained. Evaluation uses CV cohesion score and PyLDAvis visualization. Out of 10 topics, topic 5 has the highest coherence score (0.396449). PyLDAvis showed an there are 2 clusters of topics that overlap, namely topics 3 and 4. These results indicate that clusters that overlap can be combined into 1 topic cluster. Therefore, the topics that can be analyzed are 4 topics as visualized in figure 2.

Fig. 2. LDA Topic Cluster Visualization

Table 1. Topic Interpretation

Topic	Review Sentence	Topic Interpretation
1	*pertama* ***fitur tranfers*** *tidak ada setelah di uninstall baru muncul* ***fitur tranfers*** *saat mau melakukan tranfers tidak bisa alasan error bagaimana ini ovo*	Features
2	*sangat bagus dan* ***pelayanan pengaduan*** *nya selalu* ***respon*** *dengan cepat*	Service
3	*sering error kalau memang karena sinyal tapi aplikasi yang lain saja bisa di* ***akses*** *sedangkan ovo susah banget untuk* ***loginnya***	Access
4	*menurut saya digital e-wallet paling* ***aman*** *itu adalah ovo* ***keamanan*** *lumayan berlapis*	Security

Table 1 displays example sentences discussing specific topics, with bolded keywords indicating key elements for topic interpretation. The topic modeling identified four main topics: features, services, access, and security. These aspects will be used in the next stage for aspect-based sentiment analysis.

4.4 Sentiment Classification

After getting the topic that has been determined by LDA. Next, aspect labeling or aspect extraction is carried out on the reviews. Each aspect has certain words that indicate that the review discusses the intended aspect.

Based on the results of the categorization of review data aspects, it is found that out of a total of 43,452 review data, there are 24,172 reviews that cover the four aspects, while the remaining 19,280 reviews do not discuss the four aspects. Reviews are classified into positive, negative, and neutral sentiments using sentistrength. As the number of neutral sentiments dominates and lacks relevance, they are removed to avoid data imbalance. After removal, the study has 22,054 reviews, decreased by 21,398.

4.5 Data Transformation

At this stage, vectorization is performed to add weight to each word in the text document using the TF-IDF technique. The TF-IDF method is used to assign a weight to each word in a text document based on its uniqueness. In other words, the TF-IDF method helps to capture the association between words, text documents, and certain categories [24].

4.6 Evaluation

The C4.5 Decision Tree model is evaluated using three data division ratios: 70:30, 75:25, and 80:20. The ratio was taken based on the results of research which obtained high accuracy results from these three ratios [25]. The evaluation aims to measure model performance in predicting sentiment with different train-test data divisions using the confusion matrix. From the data split ratio evaluation results, the model showed excellent accuracy on both training and test data. The model has a consistent performance in correctly classifying reviews for different aspects. The 75:25 ratio resulted in the highest accuracy of 97.96%.

Table 2. Classification Report of each Aspect

Aspect	Classification	Precision	Recall	F1-Score	Accuracy
Feature	Positive	0,97	0,96	0,96	96,79%
	Negative	0,97	0,98	0,97	
Access	Positive	0,92	0,80	0,86	94,95%
	Negative	0,95	0,98	0,97	
Service	Positive	0,85	0,82	0,83	92,19%
	Negative	0,94	0,96	0,95	
Security	Positive	0,97	0,98	0,98	96,36%
	Negative	0,93	0,90	0,91	

Based on Table 2, the classification model in this study has a good accuracy rate with an average accuracy of 95.07%. Feature aspect got the highest accuracy with 96.79%. followed by Security with 96.36%, although Service and Access also got good accuracy, but there was a slight decrease in performance in identifying the Negative class. Overall, the model performed well in classifying OVO reviews.

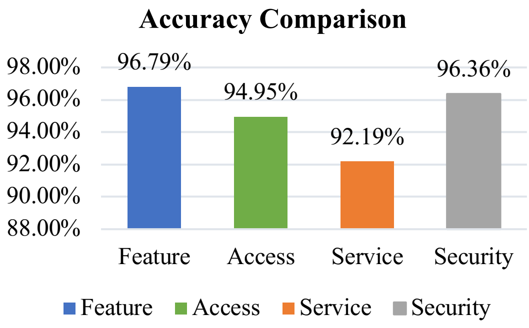

Fig 3. Accuracy Comparison of Each Aspect

Based on Figure 3, a comparison of accuracy on each aspect, it can be concluded that compared to other aspects. Overall, these results show that the C4.5 Decision Tree model performs well in predicting sentiment based on aspects with high accuracy.

5 Conclusion and Future Work

This research successfully used Latent Dirichlet Allocation (LDA) modeling topics and Decision Tree C4.5 for aspect-based sentiment analysis to identify user sentiment towards OVO services. The average accuracy obtained was 95.0725% using the optimal dataset split of 75:25. The results showed positive sentiments for Security aspects and negative sentiments for feature, access, and service aspects. The study also highlights aspects that cause technostress among users, providing valuable insights to improve OVO services in the future. To further enhance the findings, the authors recommend exploring alternative topic modeling methods such as LSA and incorporating data from multiple sources such as social media. In addition, deep learning and other machine learning models can be used for more complex sentiment analysis. OVO should focus on improving the quality of features, access, and services to reduce negative sentiment and potential technology disruption among users. Implementing these suggestions will lead to improved service quality and overall user satisfaction, ensuring OVO's continued growth and better user experience.

References

1. L. A. Abdillah, "AN OVERVIEW OF INDONESIAN FINTECH APPLICATION," *International Conference on Communication, Information Technology and Youth Study (I-CITYS)*, 2019, [Online]. Available: https://ssrn.com/abstract=3512737
2. E. Fernando, "The Influence of Perceived Risk and Trust in Adoption of FinTech Services in Indonesia," 2019.

3. AFTECH, "Annual Members Survey Asosiasi FINTECH INDONESIA 2021," pp. 1–59, 2021.
4. Boku, "Mobile Wallets Report," 2021. Accessed: Jun. 01, 2023. [Online]. Available: boku.com
5. Raihan Hasya, "Ini 10 E-Wallet yang Paling Sering Dipakai Masyarakat Indonesia Tahun 2022," *GoodStats*, 2022. https://goodstats.id/article/ini-10-e-wallet-yang-paling-sering-dipakai-masyarakat-indonesia-M4TA4 (accessed Dec. 20, 2022).
6. Y. K. Lee, "Higher innovativeness, lower technostress? comparative study of determinants on FinTech usage behavior between Korean and Chinese Gen Z consumers," *Asia Pacific Journal of Marketing and Logistics*, 2022, doi: 10.1108/APJML-05-2022-0402.
7. U. N. Ungku, A. Salmiah, M. Amin, W. Khairuzzaman, and W. Ismail, "The Impact of Technostress on Organisational Commitment among Malaysian Academic Librarians," 2009.
8. Y. K. Lee, "Impacts of digital technostress and digital technology self-efficacy on fintech usage intention of Chinese gen Z consumers," *Sustainability (Switzerland)*, vol. 13, no. 9, May 2021, doi: 10.3390/su13095077.
9. B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2012.
10. N. Aletras and M. Stevenson, "Evaluating Topic Coherence Using Distributional Semantics," *In Proceedings of the 10th international conference on computational semantics (IWCS 2013)*, 2013.
11. G. Dorfleitner, L. Hornuf, M. Schmitt, and M. Weber, "Definition of FinTech and Description of the FinTech Industry," in *FinTech in Germany*, Springer International Publishing, 2017, pp. 5–10. doi: 10.1007/978-3-319-54666-7_2.
12. BI, "PERATURAN BANK INDONESIA," 2017.
13. V. C. Nguyen and T. H. Ao, "Factors affecting consumer behaviour to use e-wallets: an empirical study from Vietnam context," *Ministry of Science and Technology, Vietnam*, vol. 64, no. 1, pp. 10–24, Apr. 2022, doi: 10.31276/VMOSTJOSSH.64(1).10-24.
14. D. M. Blei, A. Y. Ng, and Jordan I. Michael, "Latent Dirichlet Allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.
15. 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, Jun. 2019, doi: 10.1007/s11042-018-6894-4.
16. M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," 2004.
17. Y. Y. Song and Y. Lu, "Decision tree methods: applications for classification and prediction," *Shanghai Arch Psychiatry*, vol. 27, no. 2, pp. 130–135, Apr. 2015, doi: 10.11919/j.issn.1002-0829.215044.
18. J. R. Quinlan, *C4.5: Programs for Machine Learning*. CA, USA: Morgan Kaufmann, 1993.
19. W. Dai and W. Ji, "A mapreduce implementation of C4.5 decision tree algorithm," *International Journal of Database Theory and Application*, vol. 7, no. 1, pp. 49–60, Jan. 2014, doi: 10.14257/ijdta.2014.7.1.05.
20. J. G. Moreno-Torres, J. A. Saez, and F. Herrera, "Study on the impact of partition-induced dataset shift on k-fold cross-validation," *IEEE Trans Neural Netw Learn Syst*, vol. 23, no. 8, pp. 1304–1312, 2012, doi: 10.1109/TNNLS.2012.2199516.
21. M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, "Sentiment Strength Detection in Short Informal Text," *Journal of the American Society for Information Science and Technology*, vol. 61, no. 12, pp. 2544–2558, 2010.
22. M. Thelwall, K. Buckley, and G. Paltoglou, "Sentiment Strength Detection for the Social Web," *Journal of the American Society for Information Science and Technology*, vol. 63, no. 1, pp. 163–173, 2011.

23. S. P. Astuti, "Analisis Sentimen Berbasis Aspek Pada Aplikasi Tokopedia Menggunakan LDA Dan Naïve Bayes," *Bachelor's Thesis, Fakultas Sains dan Teknologi Universitas Islam Negeri Syarif Hidayatullah Jakarta*, 2020.
24. Y. T. Zhang, L. Gong, and Y. C. Wang, "Improved TF-IDF approach for text classification," *J Zhejiang Univ Sci*, vol. 6 A, no. 1, pp. 49–55, Jan. 2005, doi: 10.1631/jzus.2005.A0049.
25. K. Zahoor, N. Z. Bawany, and S. Hamid, "Sentiment analysis and classification of restaurant reviews using machine learning," *Proceedings - 2020 21st International Arab Conference on Information Technology, ACIT 2020*, Nov. 2020, doi: 10.1109/ACIT50332.2020.9300098.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (http://creativecommons.org/licenses/by-nc/4.0/), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.