IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 18 October 2023, accepted 8 November 2023, date of publication 14 November 2023, date of current version 27 November 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3332644

RESEARCH ARTICLE

The Combination of Contextualized Topic Model and MPNet for User Feedback Topic Modeling

MOHAMMAD HAMID ASNAWI[1], (Member, IEEE),
ANINDYA APRILIYANTI PRAVITASARI[1], (Member, IEEE),
TUTUT HERAWAN[2], AND **TRIYANI HENDRAWATI**[1]

[1]Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Bandung 45363, Indonesia
[2]Department of Information Systems, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur 50603, Malaysia

Corresponding author: Anindya Apriliyanti Pravitasari (anindya.apriliyanti@unpad.ac.id)

This work was sponsored by Universitas Padjadjaran and supported by the Research Center for Artificial Intelligence and Big Data, Universitas Padjadjaran.

ABSTRACT In the era of big data and ubiquitous internet connectivity, user feedback data plays a crucial role in product development and improvement. However, extracting valuable insights from the vast pool of unstructured text data found in user feedback presents significant challenges. In this paper, we propose an innovative approach to tackle this challenge by combining the Contextualized Topic Model (CTM) and the Masked and Permuted Pre-training for Language Understanding (MPNet) model. Our approach aims to create a more accurate and context-aware topic model that enhances the understanding of user experiences and opinions. To achieve this, we first search for the optimal number of topics, focusing on generating distinguishable, general, and unique topics. Next, we perform hyperparameter optimization to fine-tune the model and maximize coherence metrics. The result is an exceptionally effective model that outperforms established topic modeling methods, including LSI, NMF, LDA, HDP, NeuralLDA, ProdLDA, ETM, and the default CTM, achieving the highest coherence CV score of 0.7091. In this study, the combination of CTM and MPNet has proven highly effective in the context of user feedback topic modeling. This model excels in generating coherent, distinguishable, and highly relevant user feedback topics, capturing the nuanced nature of user feedback data. The topics generated from this model include 'Music and Audio Streaming,' 'Application Performance,' 'Banking, Financial Services, and Customer Support,' 'User Experience,' 'Other Topics,' 'Application Content,' and 'Application Features.' Our contributions include a powerful tool for developers to gain deeper insights, prioritize actions, and enhance user satisfaction by incorporating feedback into future product iterations. Furthermore, we introduce a new dataset as an open-source resource for further exploration and validation of user feedback analysis techniques and general natural language processing applications. With our proposed approach, we strive to drive business success, improve user experiences, and inform data-driven decision-making processes, ultimately benefiting both developers and users alike.

INDEX TERMS Contextualized topic model, MPNet, natural language processing, topic model, user feedback.

I. INTRODUCTION

In the big data era, which is supported by widespread internet connectivity, a huge amount of data is produced from numerous sources, providing both issues and opportunities. The abundance of data offers immense potential for gaining insights and leveraging them for various purposes. One specific type of data that is prevalent and holds significant potential is user feedback or application review data [1], [2]. In today's digital landscape, user feedback has become increasingly important in diverse domains such as e-commerce, social media, and app development. These data sources provide invaluable insights into user preferences, satisfaction levels, and, most importantly, areas for improvement [3], [4].

Understanding and harnessing the information contained from this type of data has the potential to drive business success, enhance user experiences, and help decision-making

The associate editor coordinating the review of this manuscript and approving it for publication was Biju Issac.

processes [5], [6]. However, analyzing and extracting meaningful information from the vast pool of unstructured text data found in user feedback poses significant challenges due to its unstructured nature [4], [7]. Particularly when dealing with large datasets, traditional methods struggle to handle the sheer volume and complexity of this type of unstructured data, making it difficult to process and derive valuable insights [8], [9], [10]. This is an important problem that requires addressing, as the need for fast and accurate insights to improve our product. In this study, our aim is to tackle this challenge by leveraging machine learning techniques and advanced text processing methods to create a model capable of summarizing user feedback information. By doing so, we hope to contribute to the growing body of research in natural language processing and provide businesses and developers with a powerful tool for obtaining precise insights to drive business success, enhance user experiences, inform decision-making processes, and, most importantly, identify areas for improvement.

User feedback hold substantial value for both developers and users alike, providing crucial insights into the performance, usability, and overall satisfaction levels of products and services [11]. User feedback can manifest in various forms, including ratings or scores, as well as detailed reviews themselves. In this study, we specifically focus on the textual aspect of user feedback, which offers a rich source of information. By delving into the review text, we aim to extract meaningful insights regarding user experiences, opinions, and suggestions. To facilitate this process and provide valuable insights to developers and users, we will employ machine learning techniques, specifically topic modeling, to extract the main information, also known as topics, from each feedback or review.

Topic model is a prominent unsupervised NLP technique that aims to uncover latent thematic structures within a collection of documents, making it highly relevant to our goals in this case. Extensive research has been conducted on topic modeling, employing various approaches such as statistical methods including generative probabilistic approaches like Latent Semantic Analysis (LSA) (also known as Latent Semantic Indexing (LSI)) [12] and Latent Dirichlet Allocation (LDA) [13], as well as neural network-based models like NeuralLDA [14], ProdLDA [15], and Contextualized Topic Model (CTM) [16], [17]. In our study, we will utilize the combination of CTM and the Masked and Permuted Pre-training for Language Understanding (MPNet) model [18] to enhance the topic modeling process.

CTM, which is one of the relatively recent advancements in topic modeling algorithms developed in 2021, has shown promising results in capturing contextual information and improving topic extraction accuracy. The default CTM uses SBERT as its context embedding, and while SBERT offers valuable contextual information, it has its limitations. SBERT limitations include assuming that the masked tokens are independent and predicting them individually, which is not sufficient to model the complex context dependency in natural language. Furthermore, SBERT doesn't fully leverage the position information in a sentence, resulting in position discrepancies between pre-training and fine-tuning. However, when seeking to enhance our model's performance, we turned to MPNet. MPNet, developed by Microsoft, was specifically created to address these limitations, offering robust language understanding capabilities. MPNet combines the strengths of BERT's [19] Masked Language Modeling (MLM) and XLNet's [20] Permutation Language Modeling (PLM) while overcoming their shortcomings. MPNet employs masked and permuted language modeling to capture the dependency among the predicted tokens and takes auxiliary position information as input to make the model see a full sentence, reducing position discrepancies. This innovation amplifies our model's richness and boosts efficiency, with MPNet delivering top-tier results in various applications, including sentence embedding and semantic search tasks. Moreover, MPNet offers faster processing, enhancing the practicality of our integration. The integration of CTM and MPNet allows us to extract more refined and context-aware topics from user feedback and application reviews, providing deeper insights into user experiences and opinions.

Our research endeavors are not only to obtain the best CTM-MPNet model for user feedback but also to compare its performance against existing methods, including the default CTM, which uses SBERT as the context embedding. By doing so, we aim to provide developers, researchers, and users with a powerful tool that not only extracts specific insights but also enhances the understanding of user feedback data. Our study, grounded in technical innovation and motivated by the need for more effective user feedback analysis, empowers stakeholders with actionable insights, ultimately improving the user experience in various domains.

This research makes several contributions, and the following are the main contributions of the research:

- By combining CTM and MPNet, we propose an innovative approach for user feedback topic modeling. The integration of CTM, which captures contextual information, and MPNet, which enhances language understanding, results in more accurate and context-aware topic extraction from user feedback and application reviews.
- The utilization of the proposed framework provides practical benefits for various stakeholders. For developers, it offers a deeper understanding of user experiences, enabling them to identify areas for product improvement and make data-driven decisions. For users, it enhances their satisfaction by facilitating the incorporation of their feedback into future product iterations. Additionally, researchers and analysts can leverage this framework to gain valuable insights into user sentiments
- In addition to the methodological and application contributions, this research introduces a new dataset that will soon be published as open source. The availability of this dataset will provide researchers with a valuable resource for further exploration and validation of user

feedback analysis techniques, both in unsupervised and supervised cases, as well as general NLP applications.

The structure of this paper is organized as follows: Section II provides an overview of the history of topic modeling techniques, explores the CTM, and reviews previous research that has employed CTM. In Section III, we present the proposed method, including a comprehensive overview of CTM and MPNet, as well as their combination. Section IV delves into the experimental setting, covering the dataset, data preprocessing, modeling approach, and evaluation metrics. The results of our experiments are presented and discussed in Section V, and Section VI offers the conclusion of our study along with potential directions for future research.

II. RELATED WORKS

Within the domain of topic models, researchers have developed specialized approaches to tackle diverse challenges. Classic topic models have paved the way for innovations that address issues like handling short data inputs, exploiting spatial and temporal characteristics, and understanding user preferences in specific contexts [21], [22], [23]. Our study will narrow its focus to the classic topic modeling approach, laying the foundation for a more detailed exploration of the CTM and its combination with MPNet.

CTM has emerged as a significant advancement in the realm of topic modeling techniques. CTM, developed from ProdLDA, which in turn traces its origins to LDA. ProdLDA, an extension of LDA, has demonstrated promising potential in topic modeling and has been applied to a variety of cases, including topic modeling on a TV channel's Facebook page [24] and well-known datasets such as the 20 Newsgroups dataset and Reuters Corpus Volume 1 (RCV) [14]. Moreover, researchers have explored innovative modifications of ProdLDA, such as replacing the variational autoencoder with continuous action space reinforcement learning policies, leading to a reinforcement learning approach for topic modeling [25]. ProdLDA has also found applications beyond topic modeling, including the combination of ProdLDA with convolutional neural networks (CNN) for text classification [26]. Latent Dirichlet Allocation (LDA), the foundation from which ProdLDA and CTM evolved, is widely recognized as one of the most renowned and extensively used topic modeling methods. Its impact on method development is evident from dedicated survey papers on LDA and topic modeling [27]. LDA versatility and broad applicability have established it as a popular choice for researchers investigating various domains, including geography, political science, and medical research [28], [29], [30], [31], [32], [33], [34], [35].

The base method employed in this study is the Contextualized Topic Model, developed by Bianchi et al. [16], [17]. CTM itself consists of two types: the first is the Combined Topic Model or CombinedTM type [16]. The original paper demonstrates that this approach outperforms other methods such as PLDA, MLDA, NVDM, ETM, and LDA in terms of producing more coherent and diverse topics on datasets such as Wiki20K, StackOverflow, GoogleNews, Tweets2011, and 20NewsGroups. The second type of CTM is the Zero-Shot Topic Model or ZeroShotTM [17]. The primary objective of this approach is to learn a topic model in one language and predict it for different languages or unseen documents. The research paper evaluates the quality and coherence of the topics across five languages, demonstrating that CTM with the zero-shot type performs comparably to existing models, including CTM with the Combined Topic Model type, ProdLDA, and LDA, which rely on bag-of-words representation or translations.

Despite being a relatively new method introduced in 2021, the Contextualized Topic Model has gained widespread application in both its combined and zero-shot types due to its excellent performance in topic modeling. In the case of the combined type or CombinedTM, CTM has been utilized in several studies to create topic models. For example, in a recent paper [36], the authors introduced a new family of topic coherence metrics that leverage pre-trained large language models to evaluate the interpretability and meaningfulness of topics generated by the model. CTM was compared with ETM, ATM, Top2Vec, BERTTopic, and Gibbs LDA on two datasets, namely Twitter data and the 20 NewsGroup dataset. CTM achieved the highest scores for CTCRating and CTCInstruction on the Twitter dataset and performed well on the 20 NewsGroup dataset according to CPMI and CV metrics. Another study [37] applied CTM for topic modeling on two different prostate cancer radiology reports datasets, namely RTG and JN4BD. CTM was compared with NTM and LDA, and it achieved higher coherence than both methods. Moreover, CTM successfully inferred topics related to specific anatomical regions, medical procedures, findings, diagnoses, and different modalities within these datasets. In the context of the ZeroShot type, CTM has also been used in various tasks. In a paper [38], CTM based on contextual sentence representations from multilingual models such as mBERT or XLM-R was employed. The study compared CTMs with and without fine-tuning on tasks including natural language inference, document classification, and topic classification. Additionally, CTMs were compared with continued pre-training (CPT) on the target corpus using the MLM objective, as well as with the integrated topic classification contextualized topic modeling (TCCTM) approach that modified the CTM objective with a topic classification loss. Evaluation was conducted on the English Wikipedia dataset to assess monolingual topic quality using NPMI coherence, and on aligned test documents in French, German, Portuguese, and Dutch to measure cross-lingual topic transfer using Match and KL metrics. The results demonstrated that CTMs based on contextualized representations consistently achieved higher topic coherence than LDA and ProdLDA baselines. Furthermore, fine-tuning on auxiliary tasks led to improved topic quality for CTMs. Additionally, innovative techniques have been explored in the development of CTM, as seen in a paper by González-Pizarro and Li [39]. In this study, the authors concatenated commonsense embeddings from ConceptNet Numberbatch or COMET with the SBERT embeddings used by CTM to obtain a richer representation of documents. Although the results evaluation did not demonstrate a significant increase, this approach opens up new possibilities for enhancing CTM's performance.

Furthermore, it is worth noting that numerous studies have utilized CTM in their research, including cases where CTM did not yield satisfactory results. However, upon analyzing these instances, it becomes evident that there is often a lack of recognition regarding the two different types of CTM with distinct purposes. This oversight has been a primary factor contributing to CTM's perceived underperformance in certain cases. Building on this understanding, our study focuses specifically on the CombinedTM type of CTM and combines it with MPNet, thus enhancing its performance and unlocking its full potential for extracting meaningful insights from user feedback or application reviews.

In recent years, topic modeling techniques have witnessed significant development and found applications in a wide range of fields. However, to our knowledge, there remains a notable gap: the application of topic modeling to the realm of user feedback. Despite the inherent value of user feedback data, it has been relatively underutilized. Our paper aims to bridge this gap by applying advanced topic modeling techniques to user feedback data, unlocking its potential to benefit various stakeholders, including businesses and researchers. Through this endeavor, we seek to provide a novel and valuable perspective on user feedback analysis.

III. PROPOSED METHOD

In this section, we present an innovative methodology that combines the Contextualized Topic Model with the MPNet model for the purpose of enhancing user feedback topic modeling. Our approach capitalizes on the contextualized embeddings generated by MPNet to enrich the CTM's topic modeling capabilities. To achieve this integration, we begin by providing a comprehensive overview of both the CTM and MPNet models. Subsequently, we elucidate the intricacies of their combination, the unique advantages and synergies resulting from this combination are discussed, highlighting the potential impact on uncovering meaningful topics within user feedback data.

A. CONTEXTUALIZED TOPIC MODEL (CTM)

Contextualized Topic Model was introduced in 2021 by Bianchi et al., [16], [17] offering two variants: CombinedTM [16] and ZeroShotTM [17]. Both variants utilize contextualized document embeddings from pre-trained language models, but their motivations and approaches differ. ZeroShotTM is specifically designed for cross-lingual scenarios, relying solely on pre-trained multilingual contextualized embeddings, particularly BERT models. In contrast, CombinedTM incorporates both contextualized embeddings and bag-of-words (BoW) representations as inputs, focusing on non-cross-lingual scenarios. It is essential to consider these distinctions to avoid suboptimal performance.

For our research, we exclusively concentrate on the CombinedTM variant of CTM, as our investigation does not involve cross-lingual scenarios. CTM enhances topic models by leveraging contextualized document embeddings obtained from pre-trained language models. The CTM algorithm builds upon the ProdLDA [15] neural topic model, which is grounded in variational autoencoders, by incorporating Sentence-BERT (SBERT) [40] embeddings. ProdLDA belongs to the category of models within the exponential-family PCA [41] class and shares similarities with exponential-family harmoniums [42], distinguished by the utilization of non-Gaussian priors. The core of the ProdLDA model is captured by the following equation:

$$p\left(w_n|\theta,\beta\right) \propto \prod_{k=1}^{K} p\left(w_n|z_n=k,\beta\right)^{\theta_k} \tag{1}$$

where, w_n represents the word at position n, θ is the topic proportion vector for a document, β is the topic-word matrix, and z_n is the latent topic assignment for w_n. In CTM, the SBERT embeddings are concatenated with the Bag-of-Words (BoW) representations and inputted into the inference network of ProdLDA.

CTM addresses a significant problem in traditional topic models, which typically employ BoW representations that overlook the syntactic and semantic relationships between words. By integrating contextualized representations into topic models, CTM aims to improve the coherence and diversity of the generated topics. The motivation behind the development of CTM arises from the understanding that contextualized representations from pre-trained language models possess the capability to capture both general language knowledge and corpus-specific information. Utilizing these representations can significantly enhance topic modeling by incorporating comprehensive language understanding.

B. MPNet

MPNet [18], a pre-training method for language understanding, builds upon the advancements of BERT [19] and XLNet [20] while introducing novel improvements to enhance performance. By addressing limitations found in previous pre-training objectives, such as masked language modeling (MLM) and permuted language modeling (PLM), MPNet outperforms state-of-the-art models like BERT, XLNet, and RoBERTa [43] across various natural language understanding tasks.

The motivation behind MPNet was to design a pre-training method that overcomes the challenges posed by MLM and PLM, while capitalizing on their respective advantages. While MLM fails to capture the dependency among masked tokens, PLM does not effectively utilize the full position information of a sentence.

MPNet tackles these limitations by adopting masked and permuted language modeling as its pre-training objective. The core of the MPNet method is captured in the equation 2:

$$\mathbb{E}_{z\in\mathcal{Z}_n} \sum_{t=c+1}^{n} \log P\left(x_{z_t} \mid x_{z_{<t}}, M_{z_{>c}}; \theta\right) \tag{2}$$

Equation 2 means that MPNet pre-trains the model θ by maximizing the log probability of predicting a token x_{z_t} based on its preceding tokens $x_{z_{<t}}$ and the mask tokens $M_{z_{>c}}$ in a

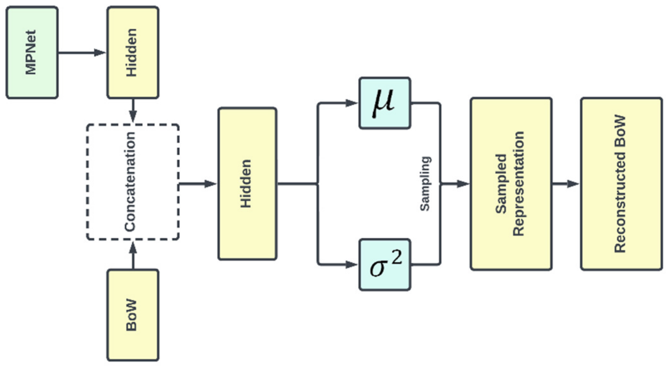

FIGURE 1. High-level scheme of the proposed approach: The combination of CTM and MPNet.

permuted sequence x_z. Here, z is a random permutation of the set $\{1, 2, \cdots, n\}$, where n is the length of the original sequence x. c is the number of non-predicted tokens in x_z, and $z_{>c}$ are the positions of the predicted tokens. $M_{z_{>c}}$ represents the mask tokens $[M]$ in positions $z_{>c}$. This formula captures the dependency among the predicted tokens and also takes the position information of all tokens as input. Refer to [18] For more details on the equation. To facilitate its enhancements, MPNet utilizes the Transformer architecture as its backbone and incorporates two key designs. Firstly, the introduction of two-stream self-attention allows the model to model the output dependency through permutation. Secondly, position compensation ensures that the model has visibility of the full sentence, including mask tokens. MPNet is pre-trained on a large-scale text corpus and subsequently fine-tuned on various downstream tasks.

C. COMBINING CTM WITH MPNET

To enhance the topic modeling process further, we propose combining CTM with MPNet by utilizing MPNet's contextualized embeddings within CTM. Building upon the strengths of both models, this integration aims to capture the contextual information and semantic representations provided by MPNet while leveraging the structure and topic modeling capabilities of CTM.

By incorporating MPNet's contextualized embeddings, we expect to improve the representation of user feedback data, allowing for a more nuanced understanding of the topics expressed within the feedback. This combination takes advantage of MPNet's ability to capture rich contextual information while also leveraging CTM's topic modeling framework.

While our model's core concept remains rooted in the principles of CTM, we have replaced SBERT with MPNet to propel our approach to new heights. This substitution is motivated by several compelling factors. First and foremost, MPNet was chosen due to its exceptional performance compared to other language models, including RoBERTa, ELECTRA, XLNet, and BERT [18]. In addition, MPNet also outperformed other language models in sentence embedding and semantic search tasks on the SBERT benchmark evaluation, boasting an impressive average performance score of 63.30. Furthermore, MPNet exhibits noteworthy speed and maintains a manageable model size, making it a practical choice for our integration. By selecting MPNet as our contextualized embedding model, we aim to tap into the wealth of contextual information it offers, aligning it seamlessly with CTM's topic modeling framework for a more holistic representation of user feedback data.

The integration of CTM with MPNet involves a series of steps, Figure 1 provides a high-level visualization of the architecture. It begins with a document as input and uses MPNet to create a contextualized embedding that captures both the meaning and context of the document. This embedding is then projected to a lower-dimensional space, with each dimension mapping to a word in the vocabulary. Concurrently, the model represents the document as a bag-of-words (BoW) vector. These two representations, the projected embedding and the BoW vector, are concatenated and passed through a hidden layer, resulting in a latent representation of the document. To approximate the posterior distribution of this latent representation, which follows a Gaussian distribution, variational inference is employed, yielding μ and σ^2 values representing the mean and variance. A vector is then sampled from this distribution, serving as the document's sampled representation. Lastly, a product-of-experts decoder is employed to reconstruct the BoW vector from the sampled representation, minimizing reconstruction loss during training.

IV. EXPERIMENTAL SETTING

In this section, we provide an overview of the experimental setup employed in our study. We begin by detailing the dataset used, including its source and collection process. Subsequently, we delve into the data preprocessing steps undertaken to prepare the input for our models. Following this, the modeling approach subsection outlines our quest for the optimal number of topics and the hyperparameter optimization process for our model. Finally, we elucidate the

TABLE 1. Five examples from the dataset.

review_id	content	score	TU_count	app_id	app_name	RC_ver
15_562	This is an amazing game! The reason I give it 4 stars is because of some reasons, first when downloading the game you don't get robux …	4	4	com.roblox.client	Roblox	2.575.424
12_945	I have to shut down the app completely every time i leave it and come back cause if i just put it in the background and go back in my …	1	1	com.hbo.hbonow	HBO Now	53.15.0.3
11_188	Keeps freezing and won't play videos or scroll. Also my paid movie playlists keep stopping instead of playing all the movies…	1	2067	com.google.android.youtube	YouTube	18.18.35
13_53	Like cash but Faster and safer! BY FAR THE BEST APP TO SEND AND RECEIVE MONEY ASAP. no nonsense, ive usrd cas …	5	4	com.squareup.cash	Cash App	3.96.0
1_404	Request I have always love this app.l ve been on it for maybe 2 or 3 years .it certainly made covid locking a lot more bearable ! Howev…	1	16	com.zhiliaoapp.musically	TikTok	29.3.2

metrics used for evaluating the performance of our integrated approach.

A. DATASET

The dataset employed in this study encompasses user feedback gathered from 15 diverse mobile applications spanning various categories. It is important to note that while we utilize this dataset for our research, we are committed to transparency and collaboration. Upon the conclusion of this project, we will release an expanded version of the dataset to the public, facilitating further research opportunities and the replication of experiments.

1) DATA GATHERING

In order to collect user feedback data for our study, we employed a data gathering approach that involved collecting application reviews or user feedback from the top 15 most downloaded apps globally. To ensure diversity in our dataset, we selected apps from multiple categories, including social media, gaming, media streaming, video editing, and payment applications. To make sure our data is current and relevant, we restricted the data to include reviews and feedback from the past five years. Therefore, the oldest data included in our dataset is from December 2018, while the most recent data is from June 2023. For each application, we gathered 1,000 user reviews, resulting in a comprehensive dataset of 15,000 data points. The following is a list of the apps we included in our data collection process:

- TikTok
- Instagram
- Facebook
- WhatsApp
- Telegram
- Zoom
- Snapchat
- Facebook Messenger
- Capcut
- Spotify
- YouTube
- HBO Max
- Cash App
- Subway Surfers
- Roblox

To gather the necessary data, we employed web scraping techniques, utilizing the google-play-scraper library in Python, to extract reviews from the Google Play Store. During the data gathering process, we ensured compliance with relevant ethical guidelines and legal regulations regarding data collection and privacy. We took steps to anonymize and aggregate the data to maintain the confidentiality and privacy of the app users whose reviews were included in our dataset.

To ensure that the collected reviews were informative and substantial, we relied on the most relevant reviews filter feature available on the Google Play Store. This feature automatically displays the reviews that are deemed most relevant to users.

TABLE 2. Dataset summary.

Column	Description
review_id	Unique identifiers for each user feedback/application review
content	User-generated feedback/review
score	Rating or star given by the user
TU_count	Number of likes/thumbs up (TU) received for the review
app_id	Unique identifier for each application
app_name	Name of the application
RC_ver	Version of the app when the review was created (RC)

2) DATA OVERVIEW

The dataset used in this study is intended to be open source and readily usable for various natural language

TABLE 3. Five examples from the dataset.

review_id	content	score	TU_count	app_id	app_name	RC_ver
15_562	This is an amazing game! The reason I give it 4 stars is because of some reasons, first when downloading the game you don't get robux …	4	4	com.roblox.client	Roblox	2.575.424
12_945	I have to shut down the app completely every time i leave it and come back cause if i just put it in the background and go back in my …	1	1	com.hbo.hbonow	HBO Now	53.15.0.3
11_188	Keeps freezing and won't play videos or scroll. Also my paid movie playlists keep stopping instead of playing all the movies…	1	2067	com.google.android.youtube	YouTube	18.18.35
13_53	Like cash but Faster and safer! BY FAR THE BEST APP TO SEND AND RECEIVE MONEY ASAP. no nonsense, ive usrd cas …	5	4	com.squareup.cash	Cash App	3.96.0
1_404	Request I have always love this app.l ve been on it for maybe 2 or 3 years .it certainly made covid locking a lot more bearable ! Howev…	1	16	com.zhiliaoapp.musically	TikTok	29.3.2

processing (NLP) projects beyond topic modeling. It includes additional columns apart from the review content, such as the score or rating, thumbs-up count, and app version. This enables the dataset to be utilized for tasks like classification and other relevant applications. In Table 2, we present a summary of the dataset, which includes a description for each column. For the 'review_id' column, each value follows a unique structure in the format of {x}_{y}, where 'x' represents a distinctive identifier for the corresponding app, and 'y' represents a unique identifier for each individual review. The 'score' column comprises integer values ranging from 1 to 5, with 1 indicating the lowest rating and 5 signifying the highest rating assigned to the app. As for the 'app_id' column, it holds identifiers assigned by the Google Play Store, serving as references to uniquely identify each app within the store's ecosystem. Additionally, Table 1 will showcase five examples from the dataset, providing a glimpse of the actual data instances.

B. DATA PREPROCESSING

In the data preprocessing stage, several steps were undertaken to ensure the data was well-prepared for subsequent modeling. The data preprocessing stages used in this study include text cleaning, tokenization, stopword removal, and lemmatization.

1) TEXT CLEANING

The text cleaning process, involves five preprocessing steps to ensure the data is in a clean and standardized format. These steps include the removal of unicode characters, lowercasing the text, removing URLs, eliminating numbers and special characters, and removing extra whitespaces. Removing unicode characters is essential to ensure that only ASCII characters remain in the user feedback data. Lowercasing the text helps to maintain consistency by converting all characters to lowercase. The removal of URLs eliminates text patterns that resemble web addresses, such as 'https' or 'www'. Additionally, the process removes numbers and special characters, retaining only letters in the text. Finally, any extra whitespaces are replaced with a single whitespace, ensuring a consistent and uniform structure throughout the data.

2) TOKENIZATION

After the initial text cleaning process, the next step in data preprocessing involves tokenization. Tokenization is a fundamental technique that breaks down the cleaned text into individual units, typically words or subwords, referred to as tokens [44]. By tokenizing the data, we transform the continuous text into a structured format that facilitates further analysis and modeling. This process involves segmenting the text into distinct tokens, enabling us to analyze the text at a granular level.

3) STOPWORD REMOVAL

After tokenization, the next step in the data preprocessing pipeline is stopword removal. Stopwords are commonly used words that do not carry significant meaning and are often removed to focus on the more informative content of the text. In our preprocessing pipeline, we leverage the widely used NLTK English stopwords dictionary for this task [45]. By applying stopword removal, we eliminate words such as 'a', 'the', 'is', and 'in' that appear frequently but do not contribute substantially to the overall meaning of the text. This process helps us filter out noise and reduce the dimensionality of the data, enabling us to focus on the more meaningful and informative words.

4) LEMMATIZATION

After stopword removal, the final step in our data preprocessing pipeline is lemmatization. Lemmatization is a process that reduces words to their base or canonical form, known as lemmas, which helps to further normalize the text. In our preprocessing pipeline, we employ the popular Spacy model, specifically 'en_core_web_sm', for performing lemmatization. By utilizing this model, we can effectively handle various linguistic nuances and accurately identify the base forms of words. Lemmatization enables us to consolidate different inflected forms of a word into a single representation, thereby reducing redundancy and enhancing the overall coherence of the data. To see the comparison before and after the preprocessing steps are applied, please refer to Table 3.

C. MODELING APPROACH

To achieve the best-performing topic model for the given user feedback data, two steps are undertaken: searching for the optimal number of topics and conducting hyperparameter optimization. The optimal number of topics serves as a critical factor in achieving topic granularity, while hyperparameter optimization using Bayesian optimization techniques further enhances the model's performance.

1) SEARCHING FOR THE OPTIMAL NUMBER OF TOPICS

In order to create a topic model that effectively captures the essence of user feedback data, it is crucial to determine the optimal number of topics. For this particular topic modeling case, the objective is to generate general and easily distinguishable topics that can facilitate focused analysis of the feedback. The aim is to avoid overly specific topics that delve into nuances of error types or app-specific details. Given the diverse nature of the feedback data, which originates from various apps and categories, the goal is to ensure that each topic represents a distinct aspect, allowing for clear differentiation between topics.

To search for the optimal number of topics, a default CTM model combined with MPNet is employed. The default parameters of the model are utilized, without modification, to ensure a fair comparison across different topic numbers. To determine the best number of topics, various performance metrics are considered. One crucial metric is topic diversity, as it aligns with the objective of creating distinguishable and general topics. The goal is to ascertain that each topic encapsulates a unique aspect of user feedback, enabling differentiation and facilitating focused analysis. By evaluating the performance metrics across the different topic numbers, we can identify the optimal number of topics that strikes a balance between topic granularity and topic differentiation.

2) HYPERPARAMETER OPTIMIZATION

In order to fine-tune the performance of our CTM combined with MPNet model, hyperparameter optimization was conducted using the classic Bayesian optimization technique that derives from Bayes theorem [46], [47], [48]. Bayesian optimization is a powerful method that efficiently explores the hyperparameter space to identify the optimal configuration for the model [47].

Bayesian optimization involves modeling the unknown objective function and iteratively updating this model based on the observed results. This allows for the intelligent selection of hyperparameters that are likely to yield better performance. Equation 3 represents the essence of Bayesian optimization.

$$x^{+} = \arg \max_{x \in A} f(x) \tag{3}$$

where x^{+} denotes the optimal set of hyperparameters, A represents the set of all possible hyperparameters, and $f(x)$ signifies the acquisition function that captures the trade-off between exploration and exploitation.

In our specific case, we optimized six hyperparameters to enhance the performance of our CTM combined with MPNet model. These hyperparameters include the activation function, dropout rate, momentum, optimization function, layer number, and neuron number. A detailed overview of each hyperparameter and its range can be found in Table 4.

To ensure comprehensive exploration of the hyperparameter space, the number of models runs followed a common rule of thumb. Each hyperparameter was multiplied by 15, resulting in a total of 90 model runs.

TABLE 4. Hyperparameter optimization space.

Hyperparameters	Space
Activation function	Categorical: sigmoid, relu, softplus, tanh, leakyrelu, rrelu, elu, selu
Dropout rate	Real: 0.0 to 0.95
Momentum	Real: 0.0 to 0.95
Optimization function	Categorical: adam, sgd
Layer number	Categorical: 1, 2, 3, 4, 5
Neuron number	Categorical: 16, 32, 64,128, 256

D. METRICS EVALUATION

To assess the performance of the topic models developed in this study, a comprehensive set of metrics has been selected to ensure a robust evaluation. Three coherence metrics, namely UCI, NPMI, and CV, will be utilized to measure the coherence and interpretability of the generated topics. Additionally, a topic diversity metric will be employed to assess the distinctiveness and generalization of the topics. This topic diversity metric plays a crucial role in the process of searching for the optimal number of topics, as discussed in the previous subsection. By utilizing this metric, we aim to create general and distinguishable topics that capture different aspects of user feedback, thereby overcoming the challenge of overly specific topics.

1) COHERENCE UCI

Coherence UCI is the first coherence measure developed by Newman et al. [49] and has since been widely used to evaluate topic models. This metric aims to capture the extent to which the top words of a topic cohesively fit together based on human judgment. It was specifically designed to correlate with the human evaluation method developed by Chang et al. [50], which marked a significant milestone in the assessment of topic models. The coherence UCI metric quantifies the coherence of a topic by measuring the pointwise mutual information (PMI) between the top words in the topic.

The PMI measures the statistical association between two words and provides an indication of how likely they co-occur within a corpus. The formula for coherence UCI is as follows:

$$UCI = \frac{2}{N \cdot (N-1)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \mathrm{PMI}\left(w_i, w_j\right) \quad (4)$$

In the equation 4, N represents the number of top words in a topic. The PMI $\left(w_i, w_j\right)$ calculates the pointwise mutual information between two words, w_i and w_j. The PMI between two words is defined as:

$$PMI\left(w_i, w_j\right) = \log \frac{P\left(w_i, w_j\right) + \varepsilon}{P\left(w_i\right) \cdot P\left(w_j\right)} \quad (5)$$

In the equation 5, $P\left(w_i\right)$ and $P\left(w_j\right)$ represent the probabilities of observing words w_i and w_j, respectively, in an external corpus. $P\left(w_i, w_j\right)$ calculates the probability of co-occurrence of two words. The term ε is a small constant added to the probabilities to avoid issues with zero probabilities.

2) COHERENCE NPMI

Coherence NPMI, introduced by Aletras and Stevenson [51], is a widely used coherence measure that leverages the concept of normalized pointwise mutual information (NPMI) to assess the co-occurrence of words within a topic. Unlike the previous coherence metrics, NPMI not only avoids zero values but also normalizes the scores between -1 and 1, offering a more comprehensive evaluation of topic coherence. The calculation of Coherence NPMI involves computing the NPMI score for each pair of top words in a topic and then averaging these pairwise scores to obtain the final coherence score for the topic. The NPMI score is determined using the following formula:

$$\mathrm{NPMI}\left(w_i, w_j\right) = \frac{\log \frac{P(w_i, w_j) + \varepsilon}{P(w_i) \cdot P(w_j)}}{-\log\left(P\left(w_i, w_j\right) + \varepsilon\right.} \quad (6)$$

3) COHERENCE CV

Coherence CV, developed by Röder et al. [52], offers a unique approach to assessing topic coherence by utilizing context vectors. These context vectors represent the co-occurrence counts of the top words within a window of five tokens in each document. Unlike the previous coherence metrics, CV does not directly use co-occurrence frequency but rather focuses on the co-occurrence of top words with other top words, allowing for greater sensitivity to the semantics of the topic. To compute coherence CV, normalized pointwise mutual information or NPMI is calculated between each word and its corresponding context vector. The NPMI values are then used to construct context vectors for all top words in the topic. For instance, let's consider a topic with top word "$word_1$". The context vector $\vec{v}_{word_1}$ may be represented as:

$$\vec{v}_{word_1} = \left[NPMI\left(word_1, word_1\right), NPMI\left(word_1, word_2\right) \ldots\right] \quad (7)$$

where $word_2$ and $word_3$ represents the second and third top word in the topic. Once the context vectors are computed, CV measures the similarity between these vectors using cosine similarity. The cosine similarity function is defined as follows:

$$\cos\left(\vec{u}, \vec{w}\right) = \frac{\sum_{i=1}^{|W|} u_i \cdot w_i}{||\vec{u}|| \cdot \parallel \vec{w}_2} \quad (8)$$

In this equation, $\vec{u}$ and $\vec{w}$ refer to the context vectors being compared, and W represents all top word context vectors. By employing cosine similarity over co-occurrence counts, CV indirectly captures the semantic similarity between words, even if they rarely co-occur directly. This allows for a more nuanced evaluation of topic coherence, as words that are semantically related but infrequently co-occurring can still contribute to the overall coherence score.

4) TOPIC DIVERSITY

Topic diversity is a relatively simple yet crucial metric used to assess the diversity of topics based on the unique words present in the topic-word distributions. Unlike the previous coherence metrics, which focus on the coherence of words within topics, topic diversity measures how distinct each topic is from the others in the model. The calculation of topic diversity involves computing the ratio of the number of unique words to the product of the number of top words per topic and the total number of topics. In essence, this metric represents the percentage of unique words in the entire set of top words across all topics.

For this specific study, topic diversity plays a significant role in the search for the optimal number of topics. As mentioned earlier, the goal is to create topics that are general and clearly differentiated from one another. Achieving 100% topic diversity is essential, as it ensures that each topic is distinct and carries unique information. The formula for topic diversity (td) is as follows:

$$td = \frac{|unique\_words|}{w \times |topics|} \quad (9)$$

Here, $|unique\_words|$ represents the count of unique words present in the topic-word distributions, w denotes the number of top words per topic, and $|topics|$ denotes the total number of topics in the topic model.

V. RESULT

In this section, we present the comprehensive results obtained from our study. We begin by discussing the outcomes of the optimal number of topics determination. Following this, we delve into the results of our hyperparameter optimization process. To assess the effectiveness of our approach, we compare model performances with existing methods, including CTM with the default SBERT model. We then delve into an exploration of the topics generated by the best model, providing insights into each distinct theme identified within the user feedback. Furthermore, a thorough discussion of the results is provided. One of the key elements of our research lies in the utilization of OCTIS [40], an advanced framework specifically designed for topic modeling tasks. Developed by Terragni et al., OCTIS offers an extensive toolkit for topic model training, analysis, comparison, and hyperparameter optimization. For those interested in replicating or examining our study, the complete code and implementation details can be accessed through the GitHub repository provided in the following link: https://github.com/mhamidasn/The-Combination-of-Contextualized-Topic-Model-and-MPNet-for-User-Feedback-Topic-Modeling

A. OPTIMAL NUMBER OF TOPICS DETERMINATION

Our main objective in determining the best number of topics is to achieve maximum topic diversity, as discussed in the previous section. We aim to create a model with distinct and general topics that facilitate focused analysis of user feedback data. To measure topic diversity, we used the topic diversity metric, which computes the ratio of unique words to the product of the number of top words per topic and the total number of topics.

In Figure 2, we present the graph displaying the topic diversity scores for different numbers of topics. As evident from the graph, the topic diversity starts at 100% for models with 5, 6, and 7 topics, showcasing that these models have highly diverse and distinguishable topics. However, as we increase the number of topics beyond 7, the topic diversity score decreases exponentially.

Based on the results from the searching for the optimal number of topics experiment, we have determined that utilizing 7 topics aligns perfectly with our research goal of achieving distinct and generalized topics in the model. Hence, for the subsequent analyses and evaluations in this study, we will use the model with 7 topics.

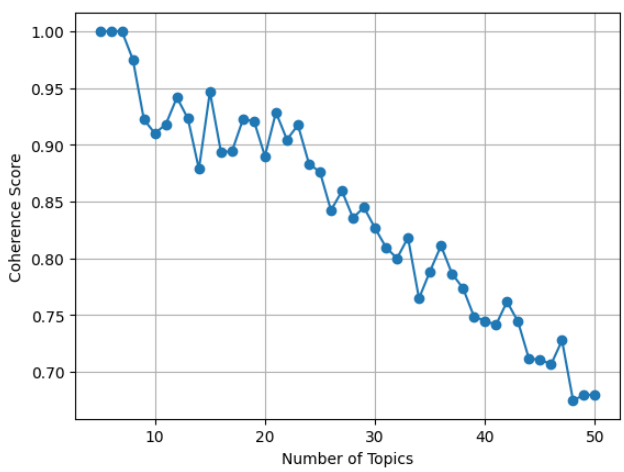

FIGURE 2. Topic diversity scores for different number of topics.

B. HYPERPARAMETER OPTIMIZATION RESULT

After determining the optimal number of topics to be 7, the next step is hyperparameter optimization to fine-tune our model for the best performance. As previously mentioned, we will optimize six hyperparameters, including activation function, dropout rate, momentum, optimization function, layer number, and neuron number. To ensure a comprehensive exploration of the hyperparameter space, we will run 90 iterations of model runs.

To gain a more comprehensive view, Table 5 presents the top 10 iterations/model runs ranked by their metrics, sorted from the highest to the lowest. Notably, iteration 88 outperformed all others, achieving the highest metric values for coherence CV, coherence UCI, and coherence NPMI, with values of 0.7091, -0.6407, and 0.0752, respectively. The next two iterations that can be considered as top-performing models after iteration 88 are 67, and 26. These iterations also exhibited promising performance, though not surpassing iteration 88.

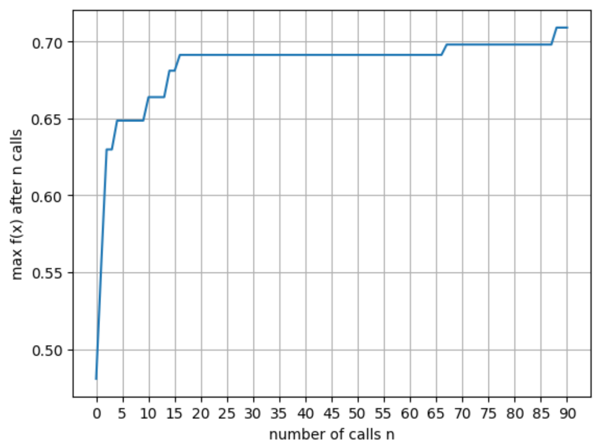

FIGURE 3. Model performance over iterations.

The hyperparameter optimization process utilized Bayesian optimization, and we tracked the convergence of the optimization using coherence CV as the evaluation metric. The convergence plot on the Figure 3 displays the performance of the model for each iteration. From the plot, it is evident that iteration 88 achieved the highest coherence CV value among all iterations.

TABLE 5. Top ten iterations/model runs ranked by each coherence metric.

Coherence CV		Coherence UCI		Coherence NPMI	
iteration	value	iteration	value	iteration	value
88	0.7091	88	-0.6407	88	0.0752
67	0.6980	74	-0.7752	26	0.0625
16	0.6913	26	-0.7862	27	0.0620
68	0.6895	27	-0.8026	67	0.0607
65	0.6821	84	-0.8247	74	0.0585
14	0.6810	16	-0.8458	68	0.0557
27	0.6807	67	-0.8767	16	0.0540
26	0.6802	37	-0.8841	84	0.0537
85	0.6801	14	-0.8996	14	0.0509
50	0.6794	68	-0.9057	60	0.0501

For the hyperparameter set used in iteration 88, the selected values were as follows: activation function = selu, dropout = 0.005, momentum = 0.02, number of layers = 1, number of neurons = 128, and optimization function = Adam. For further details on the hyperparameter sets of the top three performing models, please refer to Table 6. It provides a comprehensive breakdown of the hyperparameter configurations for the best three models, including iteration 88.

TABLE 6. Hyperparameter sets for top three performing model runs: iterations 88, 67, and 26.

Hyperparameters	It. 88	It. 67	It. 26
Activation function	Selu	Softplus	Relu
Dropout rate	0.005	0.3346	0.0901
Momentum	0.02	0.0526	0.0747
Optimization function	Adam	Adam	Adam
Layer number	1	2	3
Neuron number	128	128	16

C. MODEL COMPARISON

To assess the effectiveness of the combination of CTM with the MPNet model, our integrated approach will be evaluated against a spectrum of existing topic modeling algorithms, encompassing both conventional and neural methods. The conventional models include LSI/LSA [12], Non-Negative Matrix Factorization (NMF) [53], LDA [13], and Hierarchical Dirichlet Process (HDP) [54]. Additionally, we consider neural models such as NeuralLDA [14], ProdLDA [15], Embedded Topic Model (ETM) [55], and CTM with SBERT-base as context embeddings [16] and CTM with SBERT-large as context embeddings [16].

For a thorough assessment, our model is evaluated under two configurations: one with unoptimized hyperparameters, where the model uses the default set of hyperparameters, and another with hyperparameters optimized as discussed in the previous subsection. To ensure fairness in evaluation, all models are trained on the same preprocessed dataset, maintaining consistency in data input and processing. The primary evaluation metric used for comparing these models is coherence CV. A summary of the performance of all these models is presented in Table 7.

TABLE 7. Model comparison with existing topic modeling methods.

Model	Coherence CV
LSI	0.4589
NMF	0.3883
LDA	0.4195
HDP	0.3302
NeuralLDA	0.3314
ProdLDA	0.6074
ETM	0.4210
CTM (SBERT-base)	0.6332
CTM (SBERT-large)	0.6405
CTM (MPNet unoptimized)	0.6571
CTM (MPNet optimized)	**0.7091**

From Table 7, it becomes evident that the combination of CTM with MPNet, whether utilizing optimized or unoptimized hyperparameters, outperforms the other topic modeling methods. Specifically, CTM with MPNet (optimized) achieved the highest coherence score, registering an impressive 0.7091, signifying its remarkable ability to capture meaningful and coherent topics within user feedback data. Furthermore, CTM with MPNet (unoptimized) also delivered notable performance, with a coherence score of 0.6571, highlighting the robustness of this approach. Following closely behind, we observe that CTM with SBERT-large and CTM with SBERT-base demonstrated competitive performance with coherence scores of 0.6405 and 0.6332, respectively. In contrast, ProdLDA, while still achieving a respectable coherence score of 0.6074, falls slightly behind the CTM-MPNet variants. Based on the experiments conducted, it becomes evident that the HDP method is the least effective for our specific user feedback case, recording the lowest coherence score of 0.3302.

D. BEST MODEL TOPICS AND INSIGHTS

In the following discussion, we will delve more into the topics generated by the best model, which is the optimized CTM-MPNet model with the hyperparameter set obtained from iteration 88, exhibiting the highest performance across all four metrics. The resulting seven topics are not only distinguishable but also highly relevant and coherent.

For the topic keywords of each topic, refer to Table 8. We will now provide more detailed insights into each topic:

- Music and Audio Streaming: This topic is related to music, audio streaming, playlists, premium features, and podcasts. It seems to be strongly connected with music streaming apps, particularly the Spotify app, as users express their feedback on music-related features.
- App Performance and Troubleshooting: This topic revolves around user feedback related to app performance issues, including cache management, phone restarts, uninstallation, and dealing with crashes. Users share their experiences and challenges with app performance, making it essential for developers to address these concerns.
- Banking, Financial Services, and Customer Support: This topic focuses on user feedback related to financial transactions, customer support, account management, and issues with services like cash cards and banks. It highlights the importance of seamless and reliable financial services for app users.
- User Experience: This topic covers user feedback related to overall user experience. Interestingly, it has a strong connection with gaming experiences gained from apps like Roblox and Subway Surfers. Users share their opinions on the interface, controls, and overall usability of gaming apps.
- Other Topics: This topic captures user feedback that does not explicitly fit into other identified categories. It represents a diverse range of feedback, reflecting users'

TABLE 8. Topic keywords generated from iteration 88.

Topic	Topic Keyword
1	['listen', 'song', 'music', 'spotify', 'playlist', 'premium', 'shuffle', 'podcast', 'new', 'home']
2	['work', 'reinstall', 'open', 'cache', 'phone', 'uninstall', 'restart', 'crash', 'clear', 'fine']
3	['cash', 'card', 'account', 'support', 'customer', 'money', 'email', 'help', 'cashapp', 'bank']
4	['game', 'roblox', 'fun', 'really', 'amazing', 'good', 'lot', 'love', 'great', 'overall']
5	['pure', 'dump', 'meh', 'unnecessarily', 'aside', 'endlessly', 'scratch', 'excessively', 'disrupt', 'counter']
6	['video', 'post', 'comment', 'reel', 'watch', 'second', 'ad', 'short', 'click', 'upload']
7	['ai', 'snapchat', 'split', 'chat', 'feature', 'group', 'remove', 'option', 'pin', 'top']

opinions and suggestions outside the scope of other topics

- Content Creation within Apps: This topic focuses on user feedback related to content creation features within apps. It seems to have a strong connection with social media apps, as users share their experiences and feedback on content creation tools and functionalities.
- Application Features: This topic revolves around user feedback that talks about various application features. Users express their opinions on different features of apps, shedding light on the aspects they find useful or need improvement.

To visualize the distribution of topics generated by the best model, refer to Figure 4, which presents a bar plot of the topic frequencies. We can observe that the model performs well, as the topics are evenly generated, avoiding any strong imbalance. The most frequent topic is Topic 1, with 2922 occurrences, while the least frequent is Topic 7, with 1696 occurrences. Additionally, Table 9 provides a sample of data and its corresponding generated topic.

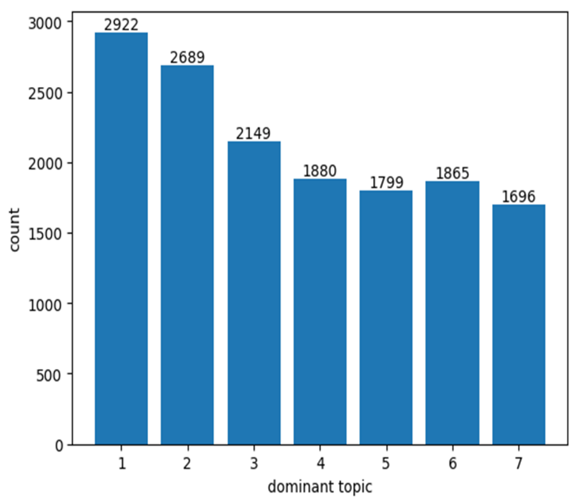

FIGURE 4. Topic distribution of the topics generated by the model with iteration 88 hyperparameter set.

E. DISCUSSION

The application of topic modeling in analyzing user feedback data holds immense importance for product developers and researchers alike. User feedback serves as a valuable source of insights, offering a direct view into user sentiments, needs, and preferences. By harnessing topic modeling techniques, developers can efficiently navigate through large volumes of user feedback and identify recurring themes, enabling them to prioritize crucial areas for improvement and innovation.

Through our innovative approach that combines CTM with MPNet, we have demonstrated that this synergy significantly enhances the topic modeling process. The resulting topics, coherent, interpretable, and effectively capturing the diversity of feedback from various users, are a testament to this improvement. Our model's performance, when compared to existing methods, both neural and conventional, becomes particularly evident. It consistently outperforms traditional methods like LSI, NMF, LDA, and HDP in producing topics with higher coherence scores that better represent user sentiments and preferences. Additionally, in comparison to neural models like NeuralLDA, ProdLDA, ETM, or even the default CTM, our CTM-MPNet integration excels at capturing the underlying structure of user feedback, resulting in more coherent and interpretable topics. Other technical advantages of our approach are evident in the balance of topic frequencies and the coherence of generated topics. The model's ability to capture subtle nuances in user feedback showcases its effectiveness in understanding the underlying semantics of user feedback. The benefit of using CTM combined with MPNet lies in its ability to leverage the contextual knowledge and language structure encoded in MPNet's pre-trained language model. This contextualized information enables the model to better understand the nuanced meanings of words, leading to more accurate and contextually appropriate topic assignments.

Our approach empowers various stakeholders by offering practical insights into user feedback data. Developers can gain a clearer understanding of the most discussed topics within user feedback, enabling them to prioritize actions and efficiently address critical user concerns. This facilitates targeted improvements and empowers developers to evaluate products or apps based on various aspects or categories. Researchers and analysts can leverage this framework to gain valuable insights into user sentiments. For instance, when seeking insights from user feedback on specific topics, such as "App Performance and Troubleshooting," our topic model can be employed to filter and analyze feedback exclusively related to these topics, facilitating more targeted

TABLE 9. Random raw data and its generated topic.

Dominant Topic	Topic Probability	Raw Input
4	0.7220643	It gets a five star because it is so much fun and I've been playing roblox since I was 7. The game has definitely changed a lot but in a good way although they did remove the oof sound): but overall it has so many different types of games and that's what makes the game unique. It makes it feel like there is a game for every type of person and there honestly is.
1	0.7784933	Great music service but I'm unable to use queue . It doesn't show any option in any song to open queue. Moreover only liked songs can be added to queue but they don't play . Swipe right ,added to queue, but queue can be find, then song not playing according to queue or the sequence I want to play them . Can't play song from each playlist the way I want it to. I have go and individually change the song after every end of a song
2	0.7727776	Consumes large amount of Data but meeting quality is poor and sometimes audio not clear or video not clear
3	0.7514825	Keeps kicking me before I can load in. Have no access to my money now and support is no help. Was fine until a day ago and now I can't ever get into my cash app. It's BS...... Tried again now a day after my first post. Still not working can't access the app to contact support and also still can't use see or takeout my money

improvements and enhancing the user experience in areas of particular concern.

In summary, our proposed approach combining CTM with MPNet provides a robust and efficient solution for topic modeling in user feedback data. The coherent and diverse topics generated enable various stakeholders to gain valuable insights, leading to targeted improvements and innovations that align with user needs and preferences. By incorporating the strengths of CTM and MPNet, we can enhance the understanding and analysis of user feedback data, ultimately driving product development and customer satisfaction.

VI. CONCLUSION

In conclusion, this study has successfully addressed the challenge of analyzing and extracting meaningful insights from vast amounts of unstructured user feedback data. By leveraging Contextualized Topic Model combined with the powerful MPNet, we have created a sophisticated model capable of summarizing user feedback information effectively.

Our experimentation led us to discover that combining CTM and MPNet outperforms existing topic modeling methods for this user feedback case. Additionally, we identified the optimal hyperparameter configuration, resulting in the best CTM model for this case. This model features the 'selu' activation function, a dropout rate of 0.005, momentum of 0.02, one layer with 128 neurons, and uses the 'adam' optimization function. The model achieved outstanding metrics with values of 0.7091, -0.6407, 0.0752, and 1.0 for coherence CV, coherence UCI, coherence NPMI, and topic diversity, respectively. These results demonstrate its ability to generate coherent, understandable, and distinguishable topics.

The topics generated by our model provide valuable insights into various aspects of user experiences and opinions. The topics include "Music and Audio Streaming", "Application Performance", "Banking, Financial Services, and Customer Support", "User Experience", "Other Topics", "Application Content", and "Application Features". These generated topics cover a wide range of crucial themes present in user feedback. Additionally, the model successfully captures context-aware and refined topics, enriching the depth of insights.

In summary, our study demonstrates the effectiveness and potential of using CTM combined with MPNet for user feedback topic modeling. By obtaining coherent and meaningful topics, developers and businesses can gain deeper understanding from user feedback data, prioritize actions, and make data-driven decisions to enhance their products and services.

Looking ahead, there are several potential future developments that can further enhance the field of user feedback analysis. Firstly, researchers can explore the usage of this dataset or variations of it, using the CTM and MPNet model, to delve deeper into specific domains or applications. Moreover, incorporating different types of topic modeling algorithms can provide comparative insights into the strengths and weaknesses of various models for different use cases.

REFERENCES

[1] B. Fu, J. Lin, L. Li, C. Faloutsos, J. Hong, and N. Sadeh, "Why people hate your App: Making sense of user feedback in a mobile App store," *Proc. Int. Conf. Knowl. Discovery Data Mining (KDD).* Chicago, IL, USA: ACM, Aug. 2013, pp. 1276–1284, doi: 10.1145/2487575.2488202.

[2] I. Morales-Ramirez, A. Perini, and R. S. S. Guizzardi, "An ontology of online user feedback in software engineering," *Appl. Ontology*, vol. 10, nos. 3–4, pp. 297–330, Dec. 2015, doi: 10.3233/ao-150150.

[3] E. C. Groen, N. Seyff, R. Ali, F. Dalpiaz, J. Doerr, E. Guzman, M. Hosseini, J. Marco, M. Oriol, A. Perini, and M. Stade, "The crowd in requirements engineering: The landscape and challenges," *IEEE Softw.*, vol. 34, no. 2, pp. 44–52, Mar. 2017, doi: 10.1109/MS.2017.33.

[4] R. I. dos Santos, E. C. Groen, and K. Villela, "An overview of user feedback classification approaches," in *Proc. REFSQ Workshops*, 2019, pp. 1–9.

[5] K. Schneider, S. Meyer, M. Peters, F. Schliephacke, J. Mörschbach, and L. Aguirre, "Feedback in context: Supporting the evolution of IT-ecosystems," in *Product-Focused Software Process Improvement*, (Lecture Notes in Computer Science), vol. 6156, M. Ali Babar, M. Vierimaa, and M. Oivo, Eds. Berlin, Germany: Springer, 2010, pp. 191–205, doi: 10.1007/978-3-642-13792-1_16.

[6] D. Pagano and W. Maalej, "User feedback in the appstore: An empirical study," in *Proc. 21st IEEE Int. Requirements Eng. Conf. (RE)*, Rio de Janeiro-RJ, Brazil, Jul. 2013, pp. 125–134, doi: 10.1109/RE.2013.6636712.

[7] E. C. Groen, J. Schowalter, S. Kopczyáska, S. Polst, and S. Alvani, ''Is there really a need for using NLP to elicit requirements? A benchmarking study to assess scalability of manual analysis,'' in *Proc. REFSQ Workshops*, 2018, pp. 1–9.

[8] B. Dresp-Langley, O. K. Ekseth, J. Fesl, S. Gohshi, M. Kurz, and H.-W. Sehring, ''Occam's Razor for big data? On detecting quality in large unstructured datasets,'' *Appl. Sci.*, vol. 9, no. 15, p. 3065, Jul. 2019, doi: 10.3390/app9153065.

[9] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, ''Deep learning for healthcare: Review, opportunities and challenges,'' *Briefings Bioinf.*, vol. 19, no. 6, pp. 1236–1246, Nov. 2018.

[10] T. B. Murdoch and A. S. Detsky, ''The inevitable application of big data to health care,'' *JAMA*, vol. 309, no. 13, pp. 1351–1352, Apr. 2013.

[11] C.-H. Hsiao, J.-J. Chang, and K.-Y. Tang, ''Exploring the influential factors in continuance usage of mobile social apps: Satisfaction, habit, and customer value perspectives,'' *Telematics Informat.*, vol. 33, no. 2, pp. 342–355, May 2016, doi: 10.1016/j.tele.2015.08.014.

[12] T. K. Landauer, P. W. Foltz, and D. Laham, ''An introduction to latent semantic analysis,'' *Discourse Process.*, vol. 25, nos. 2–3, pp. 259–284, Jan. 1998, doi: 10.1080/01638539809545028.

[13] D. M. Blei, A. Y. Ng, and M. I. Jordan, ''Latent Dirichlet allocation,'' *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.

[14] A. Srivastava and C. Sutton, ''Neural variational inference for topic models,'' in *Proc. Workshop Bayesian Deep Learn. (NIPS)*, 2016, pp. 1–12.

[15] A. Srivastava and C. Sutton, ''Autoencoding variational inference for topic models,'' 2017, *arXiv:1703.01488*.

[16] F. Bianchi, S. Terragni, and D. Hovy, ''Pre-training is a hot topic: Contextualized document embeddings improve topic coherence,'' 2020, *arXiv:2004.03974*.

[17] F. Bianchi, S. Terragni, D. Hovy, D. Nozza, and E. Fersini, ''Cross-lingual contextualized topic models with zero-shot learning,'' 2020, *arXiv:2004.07737*.

[18] K. Song, X. Tan, T. Qin, J. Lu, and T.-Y. Liu, ''MPNet: Masked and permuted pre-training for language understanding,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 16857–16867.

[19] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, ''BERT: Pre-training of deep bidirectional transformers for language understanding,'' 2018, *arXiv:1810.04805*.

[20] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, ''XLNet: Generalized autoregressive pretraining for language understanding,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 32, 2019, pp. 1–11.

[21] F. Kou, J. Du, Z. Lin, M. Liang, H. Li, L. Shi, and C. Yang, ''A semantic modeling method for social network short text based on spatial and temporal characteristics,'' *J. Comput. Sci.*, vol. 28, pp. 281–293, Sep. 2018, doi: 10.1016/j.jocs.2017.10.012.

[22] L. Shi, J. Du, M. Liang, and F. Kou, ''Dynamic topic modeling via self-aggregation for short text streams,'' *Peer-to-Peer Netw. Appl.*, vol. 12, no. 5, pp. 1403–1417, Sep. 2019, doi: 10.1007/s12083-018-0692-7.

[23] L. Shi, G. Song, G. Cheng, and X. Liu, ''A user-based aggregation topic model for understanding user's preference and intention in social network,'' *Neurocomputing*, vol. 413, pp. 1–13, Nov. 2020, doi: 10.1016/j.neucom.2020.06.099.

[24] N. Habbat, H. Anoun, and L. Hassouni, ''Extracting topics from a TV channel'S Facebook page using contextualized document embedding,'' *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XLVI-4/W5-2021, pp. 245–249, Dec. 2021, doi: 10.5194/isprs-archives-XLVI-4-W5-2021-245-2021.

[25] J. Costello and M. Reformat, ''Reinforcement learning for topic models,'' in *Proc. Findings Assoc. Comput. Linguistics (ACL)*, Toronto, ON, Canada: Association for Computational Linguistics, Jul. 2023, pp. 4332–4351. [Online]. Available: https://aclanthology.org/2023.findings-acl.265

[26] J. Ge, S. Lin, and Y. Fang, ''A text classification algorithm based on topic model and convolutional neural network,'' *J. Phys., Conf. Ser.*, vol. 1748, no. 3, Jan. 2021, Art. no. 032036, doi: 10.1088/1742-6596/1748/3/032036.

[27] H. Jelodar, Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li, and L. Zhao, ''Latent Dirichlet allocation (LDA) and topic modeling: Models, applications, a survey,'' *Multimedia Tools Appl.*, vol. 78, no. 11, pp. 15169–15211, Jun. 2019, doi: 10.1007/s11042-018-6894-4.

[28] M. Cristani, A. Perina, U. Castellani, and V. Murino, ''Geo-located image analysis using latent representations,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AK, USA, Jun. 2008, pp. 1–8, doi: 10.1109/cvpr.2008.4587390.

[29] S. Sizov, ''GeoFolk: Latent spatial semantics in web 2.0 social media,'' in *Proc. 3rd ACM Int. Conf. Web Search Data Mining*. New York, NY, USA: ACM, Feb. 2010, pp. 281–290, doi: 10.1145/1718487.1718522.

[30] D. Greene and J. P. Cross, ''Unveiling the political agenda of the European parliament plenary: A topical analysis,'' 2015, *arXiv:1505.07302*.

[31] B. Chen, L. Zhu, D. Kifer, and D. Lee, ''What is an opinion about? Exploring political standpoints using opinion scoring model,'' in *Proc. AAAI Conf. Artif. Intell.*, Jul. 2010, vol. 24, no. 1, pp. 1007–1012, doi: 10.1609/aaai.v24i1.7717.

[32] R. Cohen and D. Ruths, ''Classifying political orientation on Twitter: It's not easy!'' in *Proc. Int. AAAI Conf. Web Social Media*, vol. 7, no. 1, pp. 91–99, Aug. 2021, doi: 10.1609/icwsm.v7i1.14434.

[33] Y. Wu, M. Liu, W. J. Zheng, Z. Zhao, and H. Xu, ''Ranking gene-drug relationships in biomedical literature using latent Dirichlet allocation,'' in *Proc. Pacific Symp. Biocomputing*, 2012, pp. 422–433.

[34] Z. Jiang, X. Zhou, X. Zhang, and S. Chen, ''Using link topic model to analyze traditional Chinese medicine clinical symptom-herb regularities,'' in *Proc. IEEE 14th Int. Conf. e-Health Netw., Appl. Services (Healthcom)*, Oct. 2012, pp. 15–18.

[35] M. Paul and M. Dredze, ''You are what you tweet: Analyzing Twitter for public health,'' *Proc. Int. AAAI Conf. Web Social Media*, vol. 5, no. 1, pp. 265–272, Aug. 2021, doi: 10.1609/icwsm.v5i1.14137.

[36] H. Rahimi, J. L. Hoover, D. Mimno, H. Naacke, C. Constantin, and B. Amann, ''Contextualized topic coherence metrics,'' 2023, *arXiv:2305.14587*.

[37] T. Denti, ''Topic modeling of prostate cancer radiology reports,'' M.S. thesis, School Sci., Aalto Univ., 2022. [Online]. Available: http://urn.fi/URN:NBN:fi:aalto-202203272519

[38] A. Mueller and M. Dredze, ''Fine-tuning encoders for improved monolingual and zero-shot polylingual neural topic modeling,'' 2021, *arXiv:2104.05064*.

[39] F. González-Pizarro and R. Li, ''Contextualized topic models with commonsense knowledge,'' [Online]. Available: https://gonzalezf.github.io/projects_documents/532G_report.pdf

[40] N. Reimers and I. Gurevych, ''Sentence-BERT: Sentence embeddings using Siamese BERT-networks,'' in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, Hong Kong: Association for Computational Linguistics, 2019, pp. 3980–3990, doi: 10.18653/v1/d19-1410.

[41] M. Collins, S. Dasgupta, and R. E. Schapire, ''A generalization of principal components analysis to the exponential family,'' in *Proc. NIPS*, 2001, pp. 1–9. [Online]. Available: https://api.semanticscholar.org/CorpusID:2897627

[42] M. Welling, M. Rosen-Zvi, and G. E. Hinton, ''Exponential family harmoniums with an application to information retrieval,'' in *Proc. NIPS*, 2004, pp. 1481–1488. [Online]. Available: https://api.semanticscholar.org/CorpusID:2388827

[43] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, ''RoBERTa: A robustly optimized BERT pretraining approach,'' 2019, *arXiv:1907.11692*.

[44] G. Grefenstette, ''Tokenization,'' in *Syntactic Wordclass Tagging* (Text, Speech and Language Technology), vol. 9, H. Van Halteren, Ed. Dordrecht, The Netherlands: Springers, 1999, pp. 117–133, doi: 10.1007/978-94-015-9273-4_9.

[45] S. Bird, E. Klein, and E. Loper, *Natural Language Processing With Python: Analyzing Text With the Natural Language Toolkit*. Sebastopol, CA, USA: O'Reilly Media, 2009.

[46] J. Joyce, ''Bayes' theorem,'' in *The Stanford Encyclopedia of Philosophy*. Stanford, CA, USA: Stanford Univ., Metaphysics Research Lab, Fall 2003. [Online]. Available: https://plato.stanford.edu/archives/fall2003/entries/bayes-theorem/

[47] E. Brochu, V. M. Cora, and N. de Freitas, ''A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning,'' 2010, *arXiv:1012.2599*.

[48] W. Jia, C. Xiu-Yun, Z. Hao, X. Li-Dong, L. Hang, and D. Si-Hao, ''Hyper-parameter optimization for machine learning models based on Bayesian optimization,'' *J. Electron. Sci. Technol.*, vol. 17, no. 1, pp. 26–40, 2019.

[49] D. Newman, J. H. Lau, K. Grieser, and T. Baldwin, ''Automatic evaluation of topic coherence,'' in *Proc. Hum. Lang. Technol. Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics*, 2010, pp. 100–108.

[50] J. Chang, S. Gerrish, C. Wang, J. Boyd-Graber, and D. Blei, ''Reading tea leaves: How humans interpret topic models,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 22, 2009, pp. 1–9.

[51] N. Aletras and M. Stevenson, ''Evaluating topic coherence using distributional semantics,'' in *Proc. 10th Int. Conf. Comput. Semantics (IWCS)*, 2013, pp. 13–22.

[52] M. Röder, A. Both, and A. Hinneburg, ''Exploring the space of topic coherence measures,'' in *Proc. 8th ACM Int. Conf. Web Search Data Mining*. Shanghai China: ACM, Feb. 2015, pp. 399–408, doi: 10.1145/2684822.2685324.
[53] D. Lee and H. S. Seung, ''Algorithms for non-negative matrix factorization,'' in *Advances in Neural Information Processing Systems*, T. Leen, T. Dietterich, and V. Tresp, Eds. Cambridge, MA, USA: MIT Press, 2000. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2000/file/f9d1152547c0bde01830b7e8bd60024c-Paper.pdf
[54] Y. Teh, M. Jordan, M. Beal, and D. Blei, ''Sharing clusters among related groups: Hierarchical Dirichlet processes,'' in *Advances in Neural Information Processing Systems*, L. Saul, Y. Weiss, and L. Bottou, Eds. Cambridge, MA, USA: MIT Press, 2004. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2004/file/fb4ab556bc42d6f0ee0f9e24ec4d1af0-Paper.pdf
[55] A. B. Dieng, F. J. R. Ruiz, and D. M. Blei, ''Topic modeling in embedding spaces,'' *Trans. Assoc. Comput. Linguistics*, vol. 8, pp. 439–453, Dec. 2020, doi: 10.1162/tacl_a_00325.

MOHAMMAD HAMID ASNAWI (Member, IEEE) was born in Bandung, Indonesia, in 2002. He received the B.S. (Sarjana) degree in statistics from Universitas Padjadjaran, Indonesia, in 2023.

In 2022, he was a Research Assistant with the Research Center for Artificial Intelligence and Big Data, Universitas Padjadjaran. Continuing his journey in academia, he is currently an Associate Researcher with Universitas Padjadjaran. He is also actively seeking an opportunity to further his studies in fields related to artificial intelligence and machine learning. Outside of the academic sphere, he boasts extensive experience in various data-related roles, including data analyst and data scientist. His current research interests include artificial intelligence, machine learning, and deep learning, with a focus on unstructured data modeling, which involves natural language processing, computer vision, audio data modeling, and structured data modeling, such as time series data modeling.

ANINDYA APRILIYANTI PRAVITASARI (Member, IEEE) was born in Tuban, Indonesia, in 1984. She received the B.S. (Sarjana) and M.Si. degrees in statistics and the Ph.D. degree in statistics science from Institut Teknologi Sepuluh November, Indonesia, in 2006, 2008, and 2020, respectively.

She is currently a Lecturer and a Researcher with Universitas Padjadjaran. As part of her academic responsibilities, she is actively involved in teaching with the Department of Statistics, Universitas Padjadjaran. Additionally, she serves as the Director of the Research Center for Artificial Intelligence and Big Data. Her expertise and contributions to the field have earned her recognition, she provides her referee services to reputable journals, including IEEE Access and *European Radiology*. Moreover, her dedication to research has been acknowledged with honors, such as high-quality scientific articles in health and medicine from the Ministry of Research, Technology, and Higher Education Indonesia. Her research findings were also selected as one of the Top 20 National Posters in Artificial Intelligence by the Ministry, contributing to the advancement of Indonesian artificial intelligence innovation. Her current research interests include data mining, statistical machine learning, Bayesian analysis, natural language processing, and computer vision. She is deeply engaged in extensive research in the fields of big data and AI, continuously pushing the boundaries of knowledge and making a significant impact in the academic and research communities.

TUTUT HERAWAN received the Ph.D. degree in information technology from Universiti Tun Hussein Onn Malaysia, in 2010.

He was a Visiting Researcher with the Parallel and Distributed Computing Research Laboratory, Deakin University, Australia. He is currently an Associate Professor with the Department of Information Systems, Faculty of Computer Science and Information Technology, University of Malaya. He is also an Adjunct Professor with the Ambarrukmo Tourism Institute, Indonesia, and a Research Fellow with the Institute for Big Data Analytics and Artificial Intelligence, UiTM Shah Alam, Malaysia. He has more than 14 years of experience as an Academic Staff and has supervised several master's and Ph.D. students. He has also examined many master's and Ph.D. thesis. He has edited five Springer-series books and published more than 320 articles in various book chapters, international journals, and conference proceedings (with Scopus H-index of 34 and ISI H-index of 27). His mathematics Erdos number is four. His current research interests include applied mathematics in computer science, data science and big data, data engineering, information systems, decision support systems, data mining and knowledge discovery from databases, soft computing, and information technology for tourism. He has been named on the Top 2% World Scientists ranking by Stanford University and Elsevier BV, since 2019. He is the former Associate Editor of *Malaysian Journal of Computer Science* (ISI JCR with IF 0.405). He is the editor-in-chief and a member of the editorial board for many international journals. He is an active Reviewer for more than 50 international journals, including top reputable journals, such as IEEE Transactions on Fuzzy Systems, *Knowledge-Based Systems*, *Information Sciences*, and IEEE Access. He has delivered more than 50 keynote addresses, invited workshops, and seminars, and has actively served as the chair, the co-chair, a program committee member, and the co-organizer for numerous international conferences/workshops.

TRIYANI HENDRAWATI was born in Indonesia, in 1977. She received the B.S. (Sarjana) degree in mathematics from Universitas Indonesia and the M.Si. degree in statistics and the Ph.D. degree in statistics and data science from Institut Pertanian Bogor, Indonesia.

She is currently a Lecturer and a Researcher with the Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran. Besides being active in teaching, she also received several research grants from the Indonesian Ministry of Research, Technology, and Higher Education. Her current research interests include data mining, statistical machine learning, multivariate statistics, and statistical modeling. She is engaged in extensive research in the field of big data and AI, especially for image data and time series data.