Journal of Robotics and Control (JRC)
Volume 5, Issue 2, 2024
ISSN: 2715-5072, DOI: 10.18196/jrc.v5i2.21419

Optimizing Latent Space Representation for Tourism Insights: A Metaheuristic Approach

Thinzar Aung Win [1], Khamron Sunat [2]
[1, 2] Department of Computer Science, College of Computing, Khon Kaen University, Khon Kaen, Thailand
Email: [1] thinzar.w@kkumail.com, [2] skhamron@kku.ac.th
*Corresponding Author

***Abstract*—In the modern digital era, social media platforms with travel reviews significantly influence the tourism industry by providing a wealth of information on consumer preferences and behaviors. However, these textual reviews' complex and varied nature poses analytical challenges. This research employs advanced Natural Language Processing (NLP) techniques to process and analyze vast amounts of travel data efficiently, tackling the challenges posed by the diverse and detailed content in the tourism field. We have developed an innovative text clustering methodology that combines BERT's deep linguistic analysis capabilities (Bidirectional Encoder Representations from Transformers) with the thematic organization strengths of LDA (Latent Dirichlet Allocation). This hybrid model, further refined with the dimensionality reduction capabilities of ELM-AE and the optimization precision of PPSO (Phasor Particle Swarm Optimization), yields concise, contextually enriched text representations. Such refined data representations enhance the accuracy of K-means clustering, facilitating nuanced topic identification within the complex domain of travel reviews. This approach streamlines feature extraction and ensures rapid training and minimal loss, underscoring the model's effectiveness in distilling and reconstructing textual features. Our application of this hybrid LDA-BERT model to analyze TripAdvisor reviews of Thailand's shopping destinations reveals meaningful insights, significantly aiding in understanding customer experiences. Despite its contributions, this study acknowledges limitations, including biases in user-generated content and the intricacies of accurately interpreting sentiments and contexts within reviews. This research marks a significant step forward in utilizing NLP for tourism industry analysis, providing a pathway for future investigations to build upon.**

Keywords*—Natural Language Processing; Topic Clustering; Tourism; LDA; BERT; Embedding; Phasor Particle Swarm Optimization; ELM-Autoencoder; Feature Extraction.

I. INTRODUCTION

Social media, a category of Internet-based applications based on the Internet, has transformed how people interact and build relationships online. In the tourism sector, social media platforms have revolutionized travel planning, with a growing number of leisure travelers turning to these digital networks for insights, recommendations, and reviews, reshaping their approach to selecting destinations and organizing trips. Therefore, the travel sector must recognize the ongoing impact of technological advancements and shifts in consumer behavior on the dissemination and accessibility of travel-related information [1][2]. Similarly, in the evolution of e-tourism [3], sophisticated technologies such as artificial intelligence and big data analytics are employed to enhance the customization, efficiency, and interactivity of travel experiences. The overall travel experience could be improved through, for example, personalized travel recommendation systems based on user preferences, real-time information on attractions and transport, and interactive guides and maps on smartphones. Thus, the tourism industry increasingly depends on data for the operations and decision-making processes [4], especially as advancements in such data-driven technologies lead to rapidly expanding and frequently updated datasets.

Text data from diverse sources such as comments and blogs presents a complex and high-dimensional challenge yet contains valuable insights. Developing advanced analytical tools is crucial for efficiently mining this data, sifting through the noise to uncover relevant content and identify emerging trends. Despite the complexity of analyzing such vast, unstructured textual content, recent advancements have harnessed user-generated content to develop sophisticated recommendation systems. These systems utilize machine learning algorithms and topic modeling to customize travel and dining experiences, significantly improving personalized recommendations and bringing attention to lesser-known venues [5]. However, a gap remains in the literature regarding exploring efficient semantically based text representation methods to leverage the potential of advanced text analysis techniques fully. Thus, our study investigates contextualized text representations refined through a deep learning algorithm whose hyperparameters are fine-tuned using nature-inspired algorithms, aiming for improved topic modeling and clustering accuracy.

In the initial phase of text processing, determining the optimal approach for designing adaptive data representation methods tailored to specific tasks is paramount [6]. The extraction of valuable knowledge from textual data is fundamentally dependent on the efficacy of text representation, which is often compromised by the loss of semantic information [6], [7]. In response to this challenge, this study proposes adopting the Bidirectional Encoder Representations from Transformers (BERT) model. BERT, a state-of-the-art model utilizing Transformer-based architectures, revolutionizes Natural Language Processing (NLP) by pre-training deep bidirectional representations from unlabeled text to capture latent syntactic-semantic information, thereby establishing new standards for NLP tasks through contextual vector embeddings in a high-dimensional space [8]-[10]. Concurrently, LDA emerges as a prevalent technique for topic modeling, with the capability to

autonomously discern topics from a voluminous corpus of text documents [11].

Nonetheless, exclusive reliance on LDA-based topic models does not invariably yield satisfactory results. To overcome this limitation, the study integrates a BERT-LDA framework for constructing embedding vectors, synergizing the advantages of both methodologies to enhance the analytical depth of text representation. These vectors can be inputted into machine learning algorithms upon generating the data representations or utilized within specialized deep learning architectures [12].

Pre-trained word embeddings offer effective linguistic representations; however, their high-dimensional nature results in substantial memory requirements, posing challenges for practical applications, particularly on memory-constrained devices such as mobile phones [13]. Furthermore, integrating these embeddings into conventional clustering methodologies such as K-means, X-means, DBSCAN, and hierarchical clustering for downstream machine learning tasks becomes problematic, as these methods struggle with nonlinear data structures [14]. Furthermore, processing high-dimensional data requires significant computational time and storage. Moreover, due to the curse of dimensionality [15], it can also adversely affect the efficiency of subsequent data processing tasks, such as clustering and classification [16].

These challenges underscore the critical importance of employing dimensionality reduction strategies. While such a strategy presents a viable solution to alleviate the computational bottleneck, a comprehensive exploration of their efficacy remains essential. Earlier studies have explored the potential for reducing the dimensions of pre-trained embeddings by examining the effects of Principal Component Analysis (PCA) [17]–[19]. Dimensionality reduction can be achieved through two principal mechanisms: feature selection, which identifies and retains only the most relevant features, and feature extraction, which transforms and combines original features into a new, lower-dimensional space [20]. This study emphasizes employing feature extraction methodologies utilizing neural network architectures, given their proficiency in nonlinear unsupervised feature learning [21]. Neural network-based approaches are adept at discerning shallow and deep nonlinear patterns within data. This approach is achieved through a hierarchical processing structure, where each layer progressively identifies more straightforward features, thereby facilitating subsequent layers in extracting more abstract and complex features [22], [23]. This methodology aligns with the challenges presented by high-dimensional latent representations in real-world scenarios, where labeled data is scarce, and acquiring such data is both time-consuming and resource-intensive. The significance of this approach was highlighted in 2006 by Hinton and Salakhutdinov [24], who introduced a groundbreaking deep learning methodology for data representation. Following this, developing the autoencoder (AE) model provided a means for compressed data representation. However, its reliance on gradient-based algorithms, such as backpropagation for minimizing reconstruction error, introduced substantial computational challenges [25]-[27]. The emergence of the Extreme Learning Machine (ELM) offered a promising alternative to the conventional backpropagation method, enhancing the efficiency of the training process [28]-[30]. The advent of the ELM Autoencoder (ELM-AE) by Kasun et al. in [31] marked a notable advancement in this domain, streamlining the training process significantly and thus gaining widespread acceptance for efficient data representation. Innovations such as the generalized ELM-AE [23] and the kernel function-based ELM-AE [32] have further optimized this technique, focusing on improving optimization and computational efficiency. In light of these developments, integrating ELM-AE with hyperparameter optimization techniques has become a focal point of research, especially for training models where the selection of optimal parameters is paramount. Metaheuristic algorithms have been effectively employed to estimate network parameters, providing robust solutions to complex optimization problems [33].

Inspired by these advancements, our paper presents ELM-AE integration with hyperparameter optimization as a potent method for efficient latent representation learning, aiming to overcome the computational challenges associated with traditional backpropagation. The PPSO technique is employed for the optimal determination of hyperparameters within the ELM-AE framework, explicitly focusing on selecting the penalty coefficient and quantifying hidden nodes within the model's hidden layer. PPSO represents a significant enhancement over traditional PSO methodologies. It excels in benchmark tests by improving convergence rates and precision in identifying global optima, thanks to its innovative approach of algorithmically representing control variables through phase angles, thereby transforming PSO into a more efficient, nonparametric meta-heuristic algorithm [34]. Clustering, an unsupervised method, groups related data points into similar categories, effectively organizing topics within clusters. The K-means algorithm, a widely used partition-based clustering method in topic modeling [35], [36], categorizes unlabeled datasets into distinct groups, facilitating efficient topic categorization within these clusters. In the final stage, K-means clustering is applied to the compressed latent representations to mine meaningful topics from the input text corpora effectively.

The main contributions of this paper can be summarized as follows:

1. It augments text representation by integrating a hybrid model, which merges the capabilities of BERT with LDA. This integration produces a model that proficiently comprehends the semantic and contextual intricacies embedded within the text.
2. It contributes an optimized dimensionality reduction technique that leverages the ELM-AE in conjunction with the PPSO metaheuristic algorithm to produce efficient lower-dimensional word embeddings.
3. The research applies K-means clustering to the reduced data representation of reviews to create topic clusters.

The rest of this paper is structured as follows: Section II offers an overview of relevant literature. Section III describes the proposed methodology in detail. Experimental results and

discussions are presented in Section IV. Finally, Section V concludes the paper and outlines potential avenues for future research.

II. Related Work

Although humans naturally interpret and categorize documents into topics through reading and comprehension, such a method is not feasible for computational processes. Therefore, text representation is machines' foundational step in processing and understanding text. There are diverse techniques for text representation, which have proven successful across a range of text processing applications, such as text clustering, classification, sentiment analysis, information extraction, and information retrieval. A prevalent method is the Bag of Words (BoW) model [37], which, despite its widespread use, grapples with issues of sparsity and high dimensionality and fails to capture word relationships. A network-based Bag of Words model [38] has been developed to address these shortcomings, effectively assimilating words' structural and semantic nuances and enhancing their applicability in text classification tasks. Paper [39] improved document retrieval accuracy in text-as-data research with their Expert-Informed Topic Modeling (EITM) method, which synergistically fuses LDA with domain expert insights, employing Term-Frequency Inverse Document Frequency (TF-IDF) for text representation to pinpoint documents related to defined subjects. Paper [40] suggests a novel method for categorizing literary texts by enhancing the commonly employed term frequency-inverse document frequency (TF-IDF) approach. The method of implementing a machine learning-driven search engine was introduced in the paper [41], which leveraged natural language processing techniques to improve search results in the face of increasing digital data. This method is achieved by extracting features using TF-IDF and training with various machine learning algorithms. In addressing extractive text summarization, the proposed methodology in the paper [42] adopts a Vector Space Model coupled with topic modeling to determine sentence significance, integrating semantic similarity to refine relevance evaluation.

Recent advancements in NLP have been driven by embedding techniques, which utilize neural networks to represent word semantics in low-dimensional vector spaces. The study [43] introduced a spark-enhanced neural network model for phrase embedding that enhanced query representation in biomedical literature by effectively merging multi-word units into vector representations using advanced word embedding methods. This model effectively maintained both words and phrases within the same vector space. To address the challenge of unlabeled or diversely labeled topics in Community Question Answering, paper [44] presented TCGNN, a Graph Neural Network-based Topic Clustering framework utilizing Graph Neural Networks for enhanced topic representation for enhanced topic representation and relation mapping, demonstrating superior performance in text presentation and clustering precision across diverse datasets. Paper [45] established a deep neural network method for clustering short texts that simultaneously learns feature representations and clustering assignments, effectively transforming a high-dimensional feature space into a low-dimensional one. To enhance the efficiency of cross-domain text classification and address the complexity of 2019 novel coronavirus-related texts, the study in paper [46] proposed an advanced clustering architecture that utilized a deep sparse autoencoder optimized through backpropagation. This architecture leverages word vector models and cosine similarity for feature extraction and dimensionality reduction, culminating in K-means and mean-shift classification. In addressing the extraction of chemical-protein interactions (CPIs) from biomedical texts, the study in paper [47] introduced a Deep-contextualized Stacked Bi-LSTM (DS-LSTM) model by integrating deep contextualized word representations, an entity attention mechanism, and stacked Bi-LSTMs to improve the interpretation of context and word significance in lengthy biomedical sentences. Paper [48] highlighted the effectiveness of neural embedding feature representations for document clustering and topic modeling on social media platforms like Twitter and Reddit, which demonstrated superior performance when combined with clustering methods and evaluated against robust extrinsic measures.

In recent studies, transformer-based models such as RoBERTa [49], OpenAI's GPT-3 [50], Google's T5 [51], and Microsoft's DeBERTa [52] have been shown to advance natural language processing significantly, offering context-sensitive word representations and requiring minimal annotation, due to their innovative self-supervised learning frameworks. In paper [53], the authors developed a BERT-based model that independently assesses the relevance of document passages, showing enhanced performance with longer queries compared to shorter ones. The model evaluates documents by scoring individual passages, using all passages' first, best, or cumulative scores. In paper [54], the authors proposed the B-PROP method, an evolution of PROP that leverages BERT's contextual modeling to refine the construction of representative words, yielding significant gains in ad-hoc retrieval task performance. In the domain of biomedical relation extraction, paper [55] proposed a self-supervised graph attention network that effectively leverages BERT embeddings, a novel Gumbel Tree-GRU layer, and a knowledge fusion layer to enhance relation classification, significantly outperforming existing models on DDIExtraction 2013 and ChemProt datasets. The authors of the paper [56] performed an innovative framework for early detection and categorization of emergency events in social media, utilizing a BERT-Att-BiLSTM model for initial post-detection and an unsupervised dynamical event clustering algorithm, complemented by a supervised logistical regression model, to efficiently process and cluster different types of emergency-related data from platforms. In the paper [57], the authors employed a novel neural network framework that enhances Chinese conversational topic classification by leveraging pre-trained BERT representations, fine-tuned to produce conversational embeddings, which various neural network models then use to extract sophisticated features. The authors of the study [58] highlighted the superior feature extraction capabilities of the Transformer. They presented the T-Caps model optimized through a dynamic routing algorithm to capture low-level text features, allowing for a comprehensive sentiment classification that enhances the model's ability to discern local and overall information.

Building on the "bag of words" approach, LDA models have been extensively explored in literature for their proficiency in revealing latent topics and extracting embedded semantic knowledge from extensive document collections. Paper [59] examined various topic modeling techniques, categorizing them into standard models like VSM and LDA, clustering-based approaches for short texts, self-aggregating methods, and advanced deep learning strategies to uncover hidden semantic knowledge within diverse text corpora systematically. The authors of the paper [60] proposed an innovative hybrid sentiment analysis model tailored for the tourism sector, which synergizes sentiment lexicons with machine learning techniques. This model explicitly employs the Naive Bayes (NB) algorithm and LDA for sentiment classification. To detect patterns in extreme weather phenomena and identify regions in Romania that are particularly vulnerable to the impacts of climate change, the study [61] presents a comprehensive data processing framework that integrates LDA with clustering methods like K-prototype and K-means.

Additionally, recent advancements in enhancing text representation quality have seen numerous studies adopting a hybrid approach that combines LDA with BERT, reflecting a growing trend in research methodologies. Building upon this, the authors in the paper [62] used the combined strengths of multilingual BERT embeddings and LDA to examine topic evolution and cross-language similarities in scientific literature, bypassing machine translation limitations and subjective keyword biases. The study in paper [63] explored topic modeling by pre-processing texts and applying clustering through BERT and LDA to form vector representations that facilitate the aggregation of documents into coherent topic clusters despite LDA's known limitations in handling shorter texts.

Similarly, a study [64] proposed a novel framework that leverages user reviews and ratings to profile consumer interest in smartphones, employing a combination of LDA, BERT, and a hybrid model for topic modeling on the Roman Urdu dataset. In the paper [65], the authors developed a methodology to map the evolving field of digital twins, utilizing LDA and BERTopic for topic modeling to analyze a vast array of publications and map out the evolving landscape of this multifaceted topic. The paper research [66] developed a framework that integrated a customized Farm-Haystack question-answering system with LDA and BERT embeddings, aiming to provide quick and accurate medical information retrieval for critical domains.

The surveyed literature reveals diverse text representation techniques applied to numerous text processing tasks, with dimensionality reduction occasionally integrated into deep learning frameworks, notably transformer models for feature extraction. However, there is limited research on hybrid models incorporating these concepts. Our study proposes a hybrid embedding approach that employs deep learning for feature representation, optimized through metaheuristic algorithms to refine the model's representative learning efficacy.

III. METHODOLOGY

Fig. 1 illustrates the three-stage process of our proposed system, encompassing data collection and preprocessing, dimensionality reduction, and topic clustering. Initially, raw data is collected and preprocessed to construct text embedding vectors. Subsequently, the dimensionality reduction phase integrates the ELM-AE with PPSO to refine these embeddings into a compressed form. In the concluding phase, the K-means algorithm clusters topics. The system's application of topic modeling necessitates web scraping [67] to procure textual data from TripAdvisor, focusing on reviews related to shopping destinations in Thailand.

A. Data Collection

An automated collection framework was implemented to explore the rising interest in consumer experiences at Thai shopping venues, targeting reviews from the 'Shopping Places' category on TripAdvisor, with a specific lens on the top ten ranked venues. The data extraction utilized the "BeautifulSoup" library, enabling precise retrieval of pertinent details such as review titles, body text, reviewer geolocations, dates of posting, and corresponding ratings. The initial aggregation produced an unstructured dataset, which was then meticulously transformed into a semi-structured format conducive to Excel, facilitating systematic file creation for each venue. Ratings were harmonized into a floating-point format, categorized from "terrible" (1) to "excellent" (5) for computational uniformity. The resultant dataset, containing 9,588 entries across eight pivotal columns, was prepared for in-depth analysis. It is imperative to acknowledge the potential for sample bias, given the dataset's derivation from a singular platform, which may not accurately reflect the full spectrum of tourist demographics. Moreover, the web scraping procedure was meticulously conducted within the bounds of ethical propriety, adhering to TripAdvisor's usage policies and maintaining the anonymity of user data.

B. Text Pre-processing

Within the preprocessing phase of this study, an extensive text processing pipeline was applied to a dataset composed of TripAdvisor reviews, emphasizing processing English-language content. Language identification techniques were meticulously executed to filter out non-English reviews, excluding a mere 10 out of 8,262 reviews identified as non-English, a proportion amounting to only 0.12% of the dataset. This selective exclusion was deemed essential to mitigate potential analytical noise and enhance the model's precision, calibrated explicitly for English text analysis.

After removing extraneous elements such as URLs, emails, punctuation, emoticons, and numerical characters, Part-of-Speech (POS) tagging was employed to systematically categorize words into grammatical categories, including nouns, verbs, adverbs, and adjectives. This step highlighted the text's grammatical structures and enabled typographical error correction with the "SymSpell" algorithm. This algorithm efficiently corrects strings to a predetermined proximity from a comprehensive reference list, thus improving spelling accuracy.

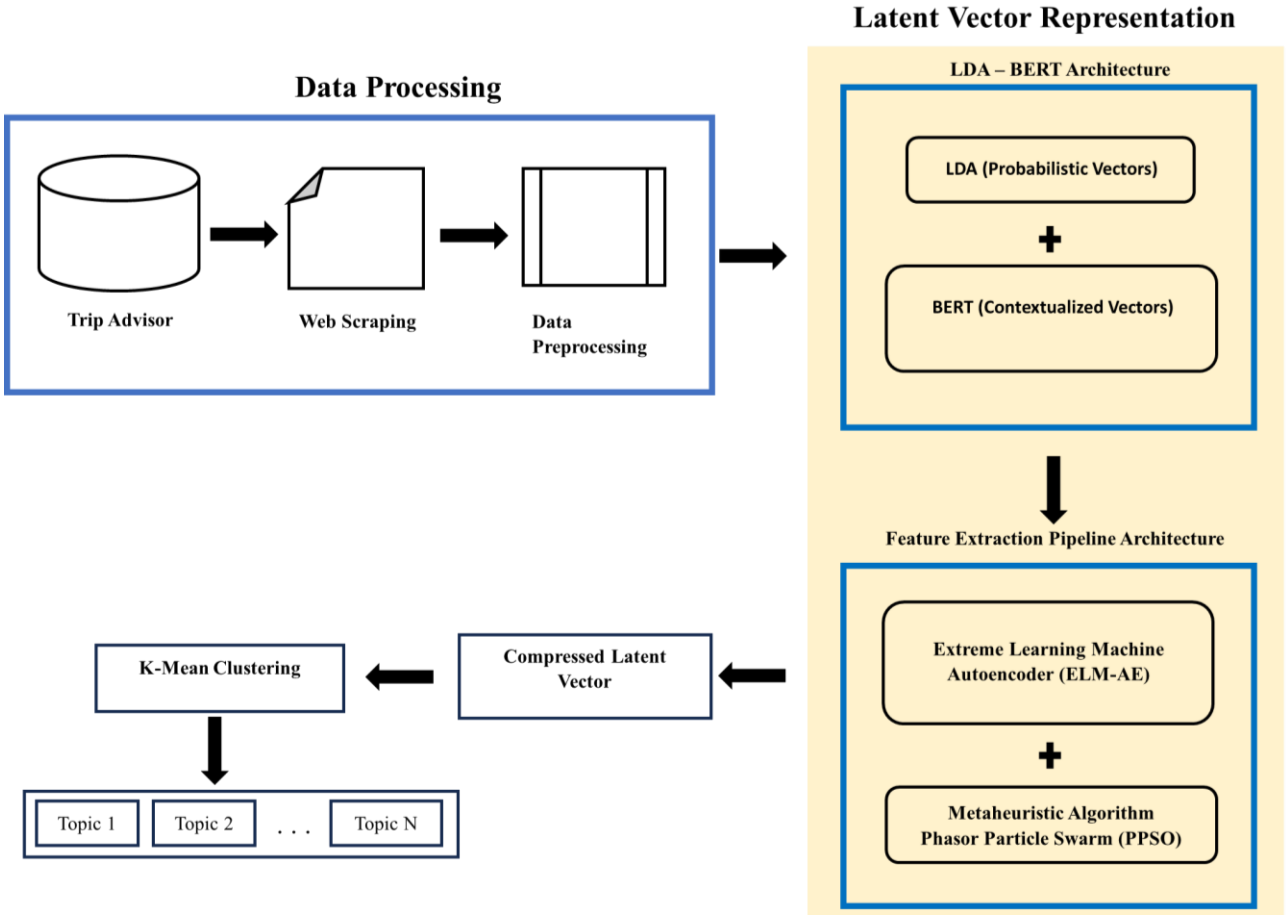

Fig. 1. Framework of the proposed model

To further refine the dataset for topic modeling, words categorized under Determiners (DT), Conjunctions (CC), Prepositions (IN), Pronouns (PRP), Interjections (UH), and Modal Verbs (MD) were removed due to their minimal contribution to the substantive analytical context. This strategic exclusion, mainly focusing on nouns and verbs, aims to reduce linguistic noise and improve the clarity and relevance of the topic modeling process. Additionally, lemmatization via the WordNet Lemmatizer was implemented to achieve lexical standardization, ensuring terminological consistency within the dataset by converging variations of words to their base forms, exemplified by unifying 'running' and 'ran' to 'run.' This comprehensive approach to text preprocessing ensures the extraction of topics that are not only relevant but also accurately representative of the dataset's core thematic content, thus maintaining the integrity and linguistic consistency of our analytical endeavors.

C. LDA-BERT Architecture

This paper aims to demonstrate the effectiveness of a topic modeling framework that merges the capabilities of BERT with LDA. Our approach unfolds through three distinct stages, starting with the generation of probabilistic vectors using the LDA model, followed by the creation of bidirectional encoded vectors from the BERT model, and culminating in the combination of the vectors yielded by BERT and LDA into a unified representation.

1) LDA's Probabilistic Vectors

Topic modeling simplifies extensive document collections into a concise topical structure, aiding in text summarization and information retrieval. It operates on the principle that documents are mixtures of topics, which are mixtures of words [68]. LDA, a prominent topic modeling approach, is a generative model to uncover topics in unseen text documents [69]. The outputs generated by LDA, including identified topics, distributions of words across topics, and distributions of topics within documents, enable a comprehensive examination of the topics prevalent within the dataset [70]. The LDA model was implemented utilizing the Gensim package in Python.

Furthermore, determining the optimal number of topics is an essential step in this process. To this end, multiple LDA models, each with varying topic quantities (k), were constructed. The optimal number of topics was identified by selecting the model with the highest coherence score, indicating topic quality. The selection of a 'k' value corresponding to the cessation of rapid growth in topic coherence typically yields meaningful and interpretable topics. The coherence score serves as a valuable metric for assessing the efficacy of a topic model, with higher scores indicating greater coherence among the topics [71]. After collecting and analyzing coherence scores for each model, the one exhibiting the highest coherence score was chosen. Consequently, the number of topics was set at eight, based on the highest coherence score achieved, indicating the most

coherent and interpretable topic distribution within the corpus.

However, LDA inherently disregards textual data's sequential and contextual nuances by adopting a bag-of-words paradigm, focusing instead on the probabilistic distribution of terms across documents to deduce thematic compositions. To address this limitation, this study proposes integrating the BERT model, aiming to capture the contextual semantic nuances of the text, thereby compensating for the deficiencies presented by the bag-of-words model.

2) Bidirectional Encoded Vectors

Initial explorations into transfer learning for natural language processing predominantly utilized recurrent neural networks. However, the prevailing trend has shifted towards employing models grounded in the Transformer architecture. Initially demonstrated to excel in machine translation tasks, the Transformer architecture has since been broadly applied across an extensive range of natural language processing applications.

BERT, an acronym for Bidirectional Encoder Representations from Transformers, represents a groundbreaking model in natural language processing (NLP) aimed at pre-training deep bidirectional representations from unlabeled textual data [72]. Subsequently, it can be fine-tuned with labeled data for diverse NLP tasks. Distinctively, BERT diverges from traditional language representation models because of its bidirectional nature, enabling it to predict words with an awareness of context from both the preceding and succeeding text. Developed as an unsupervised model, BERT leverages the vast expanse of plain text available on the web across numerous languages for training. This innovative approach empowers BERT to achieve remarkable efficacy across a spectrum of NLP applications.

The model architecture of BERT is a multi-layer bidirectional Transformer encoder derived from the original implementation published in the tensor2tensor library and detailed in [73]. Based on the following features, BERT represents a single sentence or a pair of sentences as a series of tokens. Token embeddings represent each vocabulary word with a vector using WordPiece embeddings. Position embeddings encode the positions of words within a sentence, and segment embeddings distinguish between sentences, particularly in cases of multiple sentences, by labeling words from the same sentence similarly. The input representation of a token is created by adding the segment, position, and token embeddings for that token [74]. As these input sequences are fed through the layered structure, each layer executes a sequence of operations beginning with a Self-Attention mechanism, which is then followed by a feed-forward neural network [72].

The BERT model can be used in two significant ways to pre-train deep bidirectional representations from unlabeled textual data. It can then be fine-tuned with labeled data for various NLP applications. Two variants of BERT have been introduced, namely, BERTLARGE and BERTBASE, both of which are configured for the English language. These models underwent pre-training on extensive corpora, which were meticulously compiled from various internet sources [9]. Hence, the pre-trained BERT model will be employed to generate contextual embeddings. These embeddings, typically encompassing 768 dimensions in the BERTBASE model, encapsulate the semantic essence of the entire sentence, thereby facilitating nuanced language understanding and analysis.

3) The Fusion of BERT and LDA Vectors

At this phase, the process involves merging vector outputs from both the LDA and BERT models. Specifically, the LDA model yields vectors with an eight-dimensional attribute that encapsulates thematic components, while BERT contributes with detailed sentence embeddings within a substantially broader 768-dimensional space. This systematic integration of LDA and BERT, merging LDA's proficiency in uncovering latent thematic structures in documents with BERT's adeptness at capturing textual contextual nuances, constitutes a significant methodological enhancement for analyzing tourism data. However, the resultant vectors, characterized by their high dimensionality, necessitate a transformation into a lower-dimensional latent space while preserving the contextual integrity of the vectors. This transformation is adeptly achieved by applying an ELM autoencoder. The success of this transformation also depends on the careful fine-tuning of the hyperparameters of the ELM autoencoder. Detailed elaboration of this process is provided in the subsequent section.

D. Feature Extraction Pipeline Architecture

Dimensionality reduction constitutes a pivotal process within the text mining domain [75]. It bolsters the efficacy of clustering algorithms by diminishing the dimensionality of the dataset, thereby streamlining the analysis by minimizing the number of terms processed [76]. This reduction can be accomplished via two primary strategies: feature selection and feature extraction. While feature selection involves locating and maintaining the most relevant features, feature extraction requires transforming and combining original characteristics into a new, lower-dimensional space. Prior studies have underscored the importance of such methods, especially in topic modeling within textual data, where models that generate representations in latent space, notably the Autoencoder, are frequently utilized [66], [77]. In this study, the ELM-AE is utilized to construct a feature extraction network that produces improved latent vector representations, with the system's performance assessed through topic modeling results.

The ELM-AE distinguishes itself through superior computational efficiency and adeptness at managing non-linear transformations, essential for handling the complex, high-dimensional data from LDA and BERT within the tourism domain. With a training technique that eliminates the need for iterative parameter adjustments, its ability to preserve important thematic and contextual information during dimensionality reduction makes it an excellent choice for large datasets where computational performance is critical.

1) Extreme Learning Machine Autoencoder (ELM-AE)

Building upon the ELM framework [78], ELM-AE has been introduced as a novel unsupervised learning

mechanism. The ELM-AE is trained unsupervised with an ELM algorithm, serving as a practical feature extraction strategy. The core concept behind the ELM-AE is to align the network's output as closely as possible with the intended target output, thereby ensuring a consistent transformation of the input matrix. Initially, the input is encoded by applying random weights and biases, after which the encoded data is processed through the hidden layer. This layer's output transforms an activation function, g, and is then decoded by the output weights, mapping it back into the original sample space.

A key aspect distinguishing ELM-AE [22] is its approach to determining the output weight. Contrary to the standard ELM, which minimizes the variance between the network output and the target, ELM-AE focuses on reducing the error between the product of the hidden layer output (H) and the output weight (β) and the input matrix (X). The optimal β is determined by resolving an optimization problem that minimizes the norm of β and the error between the product of H and β and the original input X, denoted as ε. The optimization is subject to constraints ensuring the orthogonality of the weight matrix.

To establish the optimal model for ELM-AE, one solves for β by minimizing the function:

$$min_\beta \frac{1}{2}\|\beta\|^2 + \frac{C}{2}\|H(W,b,X)\beta - X\|^2 = min_\beta \frac{1}{2}\|\beta\|^2 + \frac{C}{2}\|\varepsilon\|^2 \quad (1)$$

Subject to the constraints $\mathrm{W^TW = I, b^Tb = 1}$. Here, $\mathrm{H(W,b,X)} = g(W.X + b)$ denotes the hidden layer output, and g signifies the activation function applied to the hidden neurons. The subsequent output weight β is derived by solving Eq. (2), which can be represented as:

$$\beta = \left(\frac{I}{C} + H^T H\right)^{-1} H^T X \quad (2)$$

ELM-AE fuses the original high-dimensional data sample (X) with the output weight (β) to facilitate dimensionality reduction. The resulting new sample representation is computed as $\mathrm{X_{new} = X\beta^T}$, allowing the data to be represented in a lower-dimensional feature space.

2) Phasor Particle Swarm Optimization (PPSO)

PPSO is an advanced variation of the standard Particle Swarm Optimization (PSO) methodology [79]. The PPSO algorithm outperforms established PSO variants across benchmark functions. This method offers a more effective optimization solution by enhancing convergence and precision in locating global optima [34]. Recent advancements in PSO algorithms predominantly focus on selecting and refining PSO control parameters. These advancements aim to prevent premature convergence and evade local optima traps. The PPSO method [80] introduces a novel approach wherein all control variables are represented through algorithmic phase angles. This approach effectively renders the PSO as a nonparametric algorithm. However, the lack of competitive processes during evolution in PPSO can lead to stagnation in local optima, and managing a large number of particles requires sophisticated strategies to maintain diversity and computational manageability [81].

In PPSO, the selection of efficient functions for control parameters is based on periodic trigonometric functions, such as sine and cosine. Both $sine$ and $cosine$ functions, including their absolute values, oscillate between 0 to 2π radians (6.2832), adopting values between $[-1, 1]$. Their absolute values lie in the range of $[0, 1]$. These functions are inherently periodic, with sine and cosine possessing periods of 2π and π radians, respectively. The intermittent characteristics of these trigonometric functions are leveraged to represent all PSO control parameters using phase angles θ, transforming them into an θ-based function to devise diverse strategic approaches.

The phase angle θ_i of ith particle plays a critical role in guiding the particle's velocity and direction, thus impacting the overall optimization process. The Specialized Control Functions $\mathrm{p}(\theta_i^{iter})$ and $\mathrm{g}(\theta_i^{iter})$, designed to handle the particle's phase angle θ_i at each iteration, are employed. By modulating these control functions through phase angles, PPSO dynamically alters search strategies. This adaptability is crucial for effectively balancing exploration and exploitation within complex search landscapes, thereby augmenting the likelihood of uncovering optimal solutions and mitigating the risks of premature convergence.

In PPSO, the initial population of N particles is systematically dispersed within a D-dimensional problem space. Each particle is characterized by its unique phase angle θ_i, distributed uniformly within the range of $[0, 2\pi]$. This initial phase angle is pivotal, as it influences the particle's trajectory and velocity within the search space.

During each iteration, a particle's velocity vector V_i^{iter} is updated through a function of its phase angle, encompassing the particle's personal best position $P_{best,i}^{iter}$ and the global best position $Gbest^{iter}$, using the following equation:

$$V_i^{iter} = \mathrm{p}(\theta_i^{iter}) \times (P_{best,i}^{iter} - X_i^{iter}) + \mathrm{g}(\theta_i^{iter}) \times (Gbest^{iter} - X_i^{iter}) \quad (3)$$

Where $\mathrm{p}(\theta_i^{iter}) = \left|\cos(\theta_i^{iter})\right|^{2*\sin(\theta_i^{iter})}$ and $\mathrm{g}(\theta_i^{iter}) = \left|\sin(\theta_i^{iter})\right|^{2*\sin(\theta_i^{iter})}$ are the behaviors of functions. Subsequently, the particle's position is updated by integrating the newly computed velocity:

$$\vec{X}_i^{iter+1} = \vec{X}_i^{iter} + \vec{V}_i^{iter} \quad (4)$$

To further refine the search in subsequent iterations, both the phase angle and the maximum velocity limit of the particles are recalculated according to the following formulations:

$$\theta_i^{iter+1} = \theta_i^{iter} + T(\theta) \times (2\pi) \quad (5)$$

Where $T(\theta)$ is a function of $\cos(\theta_i^{iter})$ and $\sin(\theta_i^{iter})$.

$$V_{i,max}^{iter+1} = W(\theta) \times (X_{max} - X_{min}) \quad (6)$$

Where $W(\theta)$ is a function of $\left|\cos\left(\theta_i^{iter}\right)\right|^2$. These recalculations are essential as they allow the particles to adapt their search patterns by oscillating between exploration and exploitation behaviors.

3) Representation Learning by ELM Auto-Encoder with Hyperparameters Optimization

Training a learning model involves carefully selecting hyperparameters to achieve minimal prediction error. This process entails minimizing the cost function by experimenting with various hyperparameter combinations using the same training dataset. This approach helps in developing an optimal model architecture that outperforms other configurations. However, this method is time-consuming and computationally demanding [82]. Further complicating this method is the challenge of accurately identifying the optimal hyperparameters amidst a vast and complex search space, necessitating extensive and systematic exploration that may not guarantee the discovery of the most effective configuration. Metaheuristic algorithms [33] are valuable in overcoming these challenges. Their primary purpose is to solve complex optimization problems efficiently, significantly reducing errors and yielding highly accurate results. Likewise, another study [83] has concentrated on employing metaheuristic approaches to identify suitable values for the perplexity parameter in t-SNE and enhance the visualization of word embeddings.

In this context, it is crucial to carefully choose hyperparameters, such as the number of nodes in the hidden layer and the penalty coefficient in the algorithm, to construct the optimal network. We have implemented a novel technique combining PPSO with the proposed ELM-AE network to achieve this. This combination aims to identify the optimal number of hidden units and the most effective penalty coefficient for the ELM-AE model, enabling it to reconstruct its input data with the lowest error.

In the initial phase of the PPSO algorithm, the particles are systematically allocated two distinct parameters that define their position within the designated search space. The first component, which determines the number of hidden units (n_h), is assigned values covering the inclusive interval from $n_{h_{min}}$ to $n_{h_{max}}$. The second component, representing the penalty coefficient C, is similarly assigned values from the continuous range. The optimization objective, defined as the fitness of the proposed ELM-AE algorithm, is the minimization of the cost function presented by $\|\varepsilon\|^2$. The effectiveness of this approach in fine-tuning the proposed ELM-AE network is evaluated by comparing it with traditional Particle Swarm Optimization (PSO) [84], [85] and Genetic Algorithm (GA) [86] techniques. This evaluation process involves multiple iterations, culminating in a stable set of results delineating the optimal values for the hidden layer nodes and the regularization parameter ascertained through each optimization technique. A comparative analysis is then conducted to determine the optimal number of nodes in the hidden layer (n_h) and a penalty coefficient (C) that effectively balances minimizing loss and ensuring efficient training durations.

It is important to note that while the PPSO approach offers a robust mechanism for hyperparameter optimization, it introduces considerations regarding computational cost. The computational expense associated with PPSO is influenced by the iterative nature of optimizing multiple hyperparameters simultaneously. Compared to traditional methods such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), and Variance Threshold, PPSO may require additional computational resources due to its advanced search capabilities and the processing of a larger parameter space. Despite these challenges, the proposed PPSO-based optimization for the ELM-AE model represents a significant advancement in hyperparameter tuning, offering a nuanced approach that balances model accuracy with computational efficiency. In the next part of the process, the algorithm is designed to project the input data, X, onto a n_h-dimensional representation, expressed as $X_{new} = X\beta^T$. A detailed explanation of the proposed model approach is provided in Algorithm 1.

Algorithm 1: PPSO for Optimizing ELM-AE Hyperparameters

Inputs:
- $Data\ \{X\} = \{x_i\}_{i=1}^N$
- The number of hidden neurons: n_h
- The penalty coefficient: C
- The range for hidden neurons: $[n_{h_{min}}, n_{h_{max,}}]$
- The range of penalty coefficients: $[C_{min}, C_{max}]$
- The number of particles: $npop$
- The maximum iterations: $maxit$

Output:
- The Optimized number of hidden neurons $n_{h_{opt}}$
- The Optimized penalty coefficient C_{opt}
- The Optimized compressed ELM-AE representation X_{new}

Step 1: Initialize the PPSO

a) Set iteration count $t = 0$, initialize $npop$ particles with positions and velocities.
b) For each particle, generate a phase angle θ_i from $U(0,2\pi)$ and set initial velocity limits.

Step 2: PPSO Iterations

a) For each particle j:
   - Randomly assign the orthogonal random weights and biases of ELM-AE.
   - Calculate the matrix H and the matrix β using eq. (2).
   - Compute the fitness which is $\|\varepsilon\|^2$ with current n_{h_j} and C_j.
   - Initialize personal best $P_{best,j}$ and global best G_{best}.
b) While $t < maxit$:
   For each particle j:
   - Update personal best $P_{best,j}$ and global best G_{best}.
   - Update velocity V_j using eq. (3).
   - Update position X_j using eq. (4).
   - Update phase and θ_j and velocity limit $V_{j,max}$ using eq. (5) and eq. (6).
c) Evaluate the new fitness for updated positions.
d) Update P_{best} and G_{best} based on new fitness values.
e) Increment t .
f) Conclude Optimization and set $n_{h_{opt}}$ and C_{opt} from G_{best}

Step 2: Building the ELM-AE Model with Optimized Parameters

a) Initialize ELM-AE with $n_{h_{opt}}$ and C_{opt}.
b) Train ELM-AE to obtain the output weights β.
c) Project the input data X onto a new representation $X_{new} = X\beta^T$.

E. Topic Segmentation using the K-Mean Clustering Algorithm

In this part, a new representation of a n_h dimensional space as X_{new}, which has been obtained through the ELM-AE network for feature extraction, is utilized for clustering topics. We consider each row of X_{new} as a distinct point and systematically group the N points into K clusters by applying the K-means clustering algorithm [87]. This approach, which integrates unified clustering with BERT-LDA, demonstrates its efficacy in constructing topic modeling applications, as substantiated by the significant outcomes presented in [88] and [89]. We aim to precisely identify contextual topics using enhanced latent vector representations to circumvent the complexities presented by Big Data with a particular emphasis on business-related datasets. This procedure utilizes the scikit-learn library and Python to implement these machine-learning tools in an open-source environment.

To enhance the robustness of our clustering results, we have adopted a systematic approach to determine the optimal number of clusters (K). This approach involves utilizing methods such as the elbow method and silhouette analysis to evaluate the coherence and separation of clusters formed at different K values, thereby guiding the selection of an appropriate number of clusters that balances granularity with thematic distinctness.

IV. Experimental Results and Discussion

A. Experimental Configuration

This investigation harnesses unsupervised clustering and transformer-based models for topic modeling. This approach entails an extensive assessment of the efficacy of an optimized hybrid LDA-BERT latent vector representation for discerning topics within extensive textual datasets. Utilizing a corpus of 9,588 reviews from the 'Shopping Places' category on TripAdvisor, specifically selected from the top ten Thai shopping venues, this study uses preliminary data preprocessing to refine the dataset for analysis. Subsequent stages involve applying the preprocessed data to both LDA and BERT models, generating topic probability vectors and sentence vectors imbued with contextual semantic insights.

Dimensionality reduction is performed to address the challenge of high dimensionality, which is known to adversely affect the performance of clustering algorithms. This is achieved using ELM-AE for feature extraction, effectively consolidating original data features into a compact, lower-dimensional space. Following this, the PPSO algorithm is employed to optimize the hyperparameters of the ELM-AE network, thereby fine-tuning the dimensional reduction process. The resultant lower-dimensional data is then subjected to K-means clustering to form distinct topic groups, with the optimal number of clusters (k) determined through the Elbow Method. The experiments were conducted on a Google Colab Pro environment using Python version 3.10.12, with the support of a V100 GPU, 12.7 GB of system RAM, and 166.8 GB of available GPU RAM.

B. Optimal Hyperparameter Selection for ELM-AE: Balancing Loss and Computational Time

A comparative analysis of three distinct optimization algorithms was conducted to optimize ELM-AE for feature extraction from BERT and LDA generated vectors. The focus was to harmonize the trade-off between the reconstruction loss and computational expense during training. The number of nodes in the hidden layer (N_h) and the penalty coefficient (C) were the primary hyperparameters under consideration.

The optimization techniques employed were PPSO, PSO, and GA. Each algorithm was iterated 10 times to derive mean values for the loss, training time, and optimal hyperparameters. The findings, summarized in Table I, reveal an intricate relationship where $N_h$and C jointly dictates the model's loss and training duration.

An extensive evaluation suggests that an optimal number of hidden nodes falls within the range of 300 to 400, as graphically represented in Fig. 2, which signifies a critical equilibrium where the loss stabilizes. This indicates a plateau, suggesting that beyond this point, the addition of further nodes yields no significant decrease in loss. Simultaneously, the training time increases at a rate that is not overly burdensome. This equilibrium is not only dependent on N_h but also on the optimal setting of C, which becomes increasingly vital as N_h grows. The optimal C values observed in the results suggest a proportional relationship to N_h, where higher network complexities are counteracted by more robust regularization through more significant penalty coefficients.

TABLE I. ELM-AE Optimization Results for Different Optimization Algorithms

Method	Parameter	N_h Range					
		16-32	*32-64*	*64-128*	*128-256*	*256-512*	*512-776*
PPSO	**Avg. Loss**	0.1814	0.1611	0.1374	0.1101	**0.0794**	0.0596
	Avg. Training Time (s)	19.30	36.25	76.44	188.89	490.22	1001.74
	Avg. Optimal C	1.45e+21	1.54e+21	1.61e+21	1.18e+21	1.42e+02	1.11e+21
	Avg. Optimal N_h	32	64	128	256	512	775
PSO	**Avg. Loss**	0.1809	0.1604	0.1374	0.1098	**0.0794**	0.0595
	Avg. Training Time (s)	20.54	36.47	76.88	191.51	498.20	974.26
	Avg. Optimal C	1.37e+21	1.33e+21	1.26e+21	1.54e+21	1.23e+21	1.15e+21
	Avg. Optimal N_h	32	64	128	256	512	776
GA	**Avg. Loss**	0.1820	0.1614	0.1379	0.1107	**0.0798**	0.0594
	Avg. Training Time (s)	18.78	36.29	75.53	189.24	492.04	976.77
	Avg. Optimal C	1.26e+21	1.05e+21	1.36e+21	9.11e+20	6.81e+20	9.45e+20
	Avg. Optimal N_h	31	63	127	255	511	773

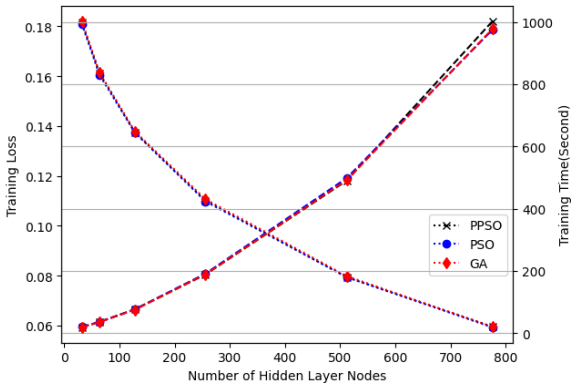

Fig. 2. ELM-AE Optimization Results for Different Optimization Algorithms

PPSO, while achieving the lowest losses, consistently suggested larger C values, indicating a rigorous penalization approach to achieve fine-grained fitting of the training data. This approach, however, comes at the cost of increased computational resources. On the other end of the spectrum, GA, with its relatively lower C values, especially in expansive N_h configurations, steers the model towards a less complex but quicker training regime, sacrificing minimal loss improvements for enhanced training efficiency. PSO's strategy balances these extremes, indicating a moderate approach to penalization and complexity. In light of these observations, it is proposed that an optimal number of hidden nodes falls within the range of 300 to 400, supported by a corresponding C value that prevents overfitting while avoiding an undue increase in training time. The carefully selected N_h-C pair produces a model that is precise in its reconstruction ability and is efficient to train.

Through this analysis, the instrumental role of metaheuristic optimization in fine-tuning ELM-AE networks has been underscored, particularly in achieving a reasonable balance between model accuracy and computational feasibility.

C. *Comparative Analysis of ELM-AE and Traditional Autoencoder*

Subsequently, an additional analysis compared the feature extraction performance between ELM-AE and a traditional autoencoder. Both models were standardized to a compressed representation dimensionality of 400. ELM-AE utilizes an ELM-based architecture and employs the ReLU activation function, with weights initialized orthogonally. Training of the ELM-AE is facilitated through a pseudo-inverse-based approach. In contrast, the traditional autoencoder is built upon a deep neural network structure, utilizes LeakyReLU for activation, and follows standard weight initialization, with backpropagation as its training method.

According to the comparative analysis, the remarkable training duration reduction to approximately 0.5485 seconds for the ELM-AE, as indicated by Table II, marks a substantial advancement in computational efficiency when contrasted with the conventional training approaches that exceed 200 seconds. Although the training loss for the ELM-AE, documented at 0.0942, exceeds that observed in the traditional autoencoder, the marginal increase is mitigated by the considerable reduction in training time. Furthermore, the validation loss for ELM-AE is 0.1160, higher than the validation loss of 0.0523 for the standard autoencoder. This suggests that there is a trade-off between the model's capacity to generalize new data and its training speed.

TABLE II. COMPARING THE FEATURE EXTRACTION PERFORMANCE OF ELM-AE WITH TRADITIONAL AUTOENCODER

Feature	ELM-AE	Traditional Autoencoder
Model Architecture	ELM-based	Deep Neural Network
The compressed representation Size	400	400
Activation Function	ReLU	LeakyReLU
Weight Initialization	Orthogonal	Standard
Training Method	Pseudo-Inverse Based	Backpropagation
Training Time (second)	**0.5485**	202.6577
Training Loss	0.0941	**0.0451**
Validation Loss	0.1160	**0.0523**

To enhance the interpretability of semantic features, we conducted a detailed qualitative analysis, assessing the most influential features within each cluster and scrutinizing the associated terms and phrases within the textual data. For example, Cluster 0, as defined by the ELM-AE model, predominantly encapsulates themes related to markets and shopping experiences, characterized by recurring terms such as “food,” “market,” “stall,” and “shopping.” Conversely, Cluster 1 predominantly features a lexicon of structured shopping venues and food courts, exemplified by terms like "mbk," "shop," "food court," and "siam." Cluster 2, also delineated by ELM-AE, signifies a shift towards more high-end retail environments, indicated by frequent references to “brand” and “store.” In contrast, the clusters formed by the traditional autoencoder present a more nuanced landscape. Cluster 0 appears to reflect narratives centered around nocturnal commerce, indicated by a constellation of terms including “market,” “food,” “shop,” and “night.” Cluster 1, while still within the retail domain, is distinguished by terms such as “mall,” “bag,” “siam,” “bangkok” as well as “boat,” “river” and “asiatique” suggesting a fusion of shopping experiences with waterside attractions. Cluster 2 diverges towards gastronomy and dining, with “restaurant,” “bangkok,” “thai” and “food” surfacing as prominent themes. The traditional autoencoder’s clusters reveal a rich tapestry of semantic connections, displaying a substantial overlap of terms within and across clusters, indicative of a granular and nuanced feature representation.

In comparison, the ELM-AE tends to generate clusters with greater cohesion, where the textual data within each cluster shares higher similarity regarding the extracted features.

Hence, while the traditional autoencoder may require more computational time for training, it potentially offers a set of features with enhanced interpretability attributed to the granularity and depth of the semantic structure within the clusters. In contrast, the ELM-AE, notable for its computational expediency and clear thematic delineation, might be more suited to applications where rapid feature extraction is paramount and a broader thematic distinction suffices.

TABLE III. COMPARISON OF TEXT REPRESENTATIONS USING K-MEANS CLUSTERING

Text Representation	Sil. Score								
	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 6$	$k = 7$	$k = 8$	$k = 9$	$k = 10$
Optimized LDA-BERT	**0.4110**	0.3048	0.2888	0.2638	0.2647	0.2488	0.2368	0.2345	0.2312
LDA-BERT	0.2154	0.0989	0.0866	0.0496	0.0488	0.0468	0.0419	0.0407	0.0401
BERT	0.2172	0.0997	0.0874	0.0504	0.0498	0.0474	0.0427	0.0412	0.0399
LDA	0.3053	0.4474	0.5521	0.6521	0.7124	0.7595	0.7951	0.7680	0.7434
TF-IDF	0.0083	0.0070	0.0072	0.0065	0.0068	0.0061	0.0064	0.0062	0.0062

D. Integrated Evaluation of Text Representation Techniques Using K-Means Clustering and Silhouette Scores

To conduct a comprehensive evaluation of our proposed text representation method, the optimized LDA–BERT hybrid model, it is essential to perform comparative analyses against a range of text representation techniques, including the standalone LDA-BERT, BERT, LDA, and TF-IDF models. The effectiveness of these text representation techniques was thoroughly examined using K-means clustering, with the number of clusters varying from two to ten ($k = 2$ to $k = 10$). The silhouette coefficient—a metric for assessing the quality of clustering—was utilized to discern the Coherence and separation of the clusters formed by each method [90]-[92]. The results, as depicted in Table III, show the comparative performance of the techniques across varying cluster sizes.

Notably, the optimized LDA-BERT model achieved superior clustering performance, with the highest silhouette score of 0.4110 at $k = 2$, and maintained a descending trend as the number of clusters increased, ending with a score of 0.2312 at $k = 10$. This performance suggests an optimal cluster formation at lower k values, with diminishing cluster definition at higher k values. In our findings, the uniformity in performance decrement may suggest an equitable contribution from our model's LDA and BERT components. This advancement underscores the effectiveness of our feature extraction approach applied to the LDA-BERT vectors, which has proven beneficial in maintaining a stable and favorable trend in silhouette scores.

Moreover, the enhanced performance suggests that integrating our optimization techniques with the hybrid LDA-BERT embeddings results in the formation of balanced and semantically more relevant clusters for topic discovery. Our findings, however, show that the original LDA-BERT representation begins with a moderate silhouette score at $k = 2$ (0.2154) and experiences a steady decrease without any notable spikes or spades, reaching its lowest at $k = 10$ (0.0401). In the case of BERT, it displayed a moderately robust initial clustering performance at $k = 2$ with a silhouette score of 0.2172. This performance mildly fluctuated as the number of clusters increased, with a silhouette score of 0.0874 at $k = 4$, slightly improving at $k = 7$ to a score of 0.0474 before ultimately descending to its lowest value of 0.0399 at $k = 10$. These results suggest that while BERT can form coherent clusters when dealing with a smaller set of clusters, its ability to delineate between clusters diminishes as the number of clusters expands. Meanwhile, the LDA model showed a progressive improvement in silhouette scores with increasing cluster sizes up to $k = 7$, suggesting that the cluster structure becomes more defined as the number of clusters increases up to a certain point. The consistently low silhouette scores for TF-IDF across all cluster sizes highlight its limitations in effectively differentiating text clusters within large datasets, suggesting that traditional text representation techniques may not be sufficient for such complex clustering tasks.

The optimized LDA-BERT model has exhibited commendable clustering performance, and silhouette score metrics have quantitatively substantiated its effectiveness. Nevertheless, it is imperative to consider the inherent constraints of the K-means clustering algorithm employed within our study. A known limitation of the K-means algorithm is its sensitivity to the initial selection of cluster centroids, which can significantly influence the outcome of the clustering process, often necessitating multiple iterations to achieve a stable and reliable configuration. To circumvent this challenge, our study drew a comparison with previous research wherein an enhanced method, Genetic Algorithm-based K-means (GA-K-Means), was utilized to refine the initialization phase of the K-means algorithm. This approach strategically employs genetic algorithms to determine optimal initial centroids, thereby addressing the random initialization limitation intrinsic to standard K-means.

As indicated in Table IV, the GA-K-Means method demonstrated a notable increase in silhouette scores, reaching 0.6606, compared to the traditional K-means score of 0.4189 documented in previous studies. However, the silhouette score of the proposed K-Means method in our study is 0.4110, which closely aligns with the earlier K-means results. This slight variance in silhouette scores could suggest that while the optimized initialization technique offers an improvement, the fundamental characteristics of the K-means algorithm continue to play a determinative role in the clustering quality, underscoring the necessity for a reasonable selection of initialization strategies in K-means-based clustering applications.

E. The Evaluation of the Optimized LDA-BERT Representation Using Two Different Clustering Algorithms

This step assessed the robustness of the optimized LDA combined with BERT representations using a clustering algorithm distinct from the K-means family. Furthermore, this cross-validation was imperative to ascertain the reliability of the word clusters generated.

TABLE IV. COMPARISON OF SILHOUETTE SCORES FOR K-MEANS CLUSTERING FROM PREVIOUS RESEARCH AND THIS STUDY

Methods	Sil. Score
K-Means (Bouabdallaoui et al., 2023) [93]	0.4189
GA-K-Means (Bouabdallaoui et al., 2023) [93]	**0.6606**
K-Means (This Study)	0.4110

It is hypothesized that if the same words are consistently grouped across different algorithms, it will substantiate the integrity of the clustering process. A critical aspect of clustering analysis is determining the optimal cluster count, K, which is instrumental in achieving superior algorithmic performance. The elbow method [94] serves as a cardinal criterion for this selection process by analyzing the within-cluster sum of squares (WCSS) [95], which quantifies the sum of the squared distances for each data point relative to its cluster centroid. In Fig. 3, the elbow curve illustrates the WCSS against the number of clusters k. Upon careful analysis of the curve, a substantial decrease in WCSS is observed as the number of clusters increases from $k = 1$ to $k = 3$, suggesting a considerable improvement in cluster tightness with each additional cluster within this range. The elbow of the curve is observed at $k = 3$, where the WCSS is approximately 0.75×10^9.

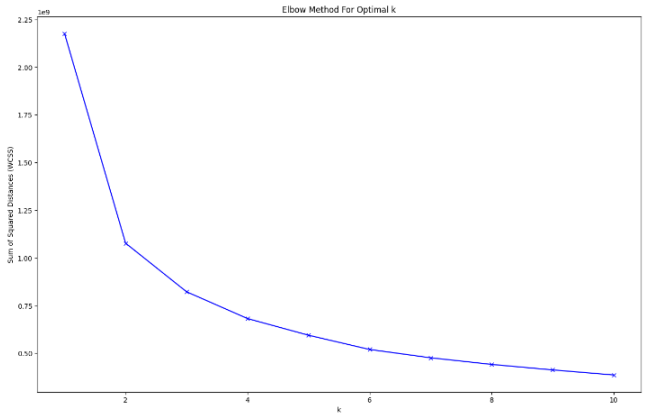

Fig. 3. Elbow Curve

Consequently, the empirical evidence from the elbow plot substantiates the adoption of $k = 3$ as the most suitable number of clusters for this analysis. However, a notable variation was observed in a subsequent iteration, indicating that $k = 4$ also provides a comparable interpretation to $k = 3$. Beyond this point, the curve begins to flatten, implying that the incremental benefit of additional clusters is marginal compared to the complexity they add. To resolve this uncertainty, our focus shifted to three specific cluster counts: $k = 3$, $k = 4$, and $k = 5$. These were evaluated using K-means and Agglomerative clustering methods [96], [97]. In addition to the silhouette coefficient, which remains a robust measure of cluster quality, a Coherence coefficient was incorporated. This new metric, based on Coherence, serves as an indicator of the interpretability of the clusters.

As depicted in Table V, when analyzing the outcomes with $k = 3$, it is observed that both K-means and Agglomerative Clustering produce similar coherence scores, with values of 0.6068 and 0.6039, respectively. Although there is a minor difference in their silhouette scores, the proximity of their coherence scores implies that both methods probably categorize words similarly. This observation suggests a consistent and reliable performance in terms of clustering from both algorithms.

In transitioning to $k = 4$, the coherence scores of both K-means and Agglomerative Clustering algorithms show close alignment, with K-means scoring 0.5891 and Agglomerative Clustering scoring 0.5948. Although the lower Silhouette Score observed in Agglomerative Clustering compared to K-means, its coherence score surpasses slightly. This observation suggests that while the clusters in Agglomerative Clustering might not be as well-separated as those in K-means, the interpretability and consistency in categorizing words in both algorithms are effectively maintained. As the cluster number reaches $k = 5$, the silhouette scores for both K-means and Agglomerative Clustering algorithms decrease to 0.2638 and 0.2618, respectively. This observation suggests that increasing the number of clusters does not invariably result in more distinctly defined clusters. The relative robustness demonstrated by the coherence scores for K-means and Agglomerative Clustering, which are 0.5865 and 0.5817, respectively, indicates that despite the increment in the number of clusters, the interpretability and meaningfulness of the clusters are effectively maintained.

The presented findings, therefore, indicate that while the K-means algorithm slightly outperforms Agglomerative Clustering in terms of Silhouette Score, both algorithms offer comparable performance in terms of coherence score across all tested cluster counts. Furthermore, these results suggest a diminishing return on cluster definition quality with increasing k beyond 3, as evidenced by the decreasing silhouette scores. Nevertheless, the relative stability of the coherence score up to $k = 5$ implies that the clusters maintain their meaningfulness and interpretability. The selection of $k = 3$ is supported by the "elbow method" observed in the elbow curve presented in the study, affirming the theoretical reasoning that this cluster count optimally balances cluster cohesiveness and algorithmic efficiency.

TABLE V. PERFORMANCE OF OPTIMIZED EXTRACT FEATURE REPRESENTATION WITH TWO DIFFERENT CLUSTERING ALGORITHMS

	Clustering Algorithm	K-means	Agglomerative Clustering
k=3	**Sil. Score**	**0.3048**	0.2653
	Coh scores	0.6068	0.6039
k=4	Sil. Score	0.2888	0.2318
	Coh scores	0.5891	0.5948
k=5	Sil. Score	0.2638	0.2618
	Coh scores	0.5865	0.5817

F. Integration of UMAP Visualization and Word Cloud Insights in Cluster Analysis

The assessment of cluster quality was performed utilizing the Silhouette Coefficient, a measure that evaluates the degree of separation between clusters. This coefficient ranges from -1 to 1, with values approaching 1 reflecting a higher degree of distinction among clusters. Subsequently, we employed Uniform Manifold Approximation and Projection (UMAP) [98] as a sophisticated method for reducing dimensionality. UMAP is grounded in a theoretical framework informed by Riemannian geometry and algebraic topology, offering an advanced alternative to other reduction techniques such as t-distributed stochastic neighbor embedding (t-SNE) and principal component analysis (PCA). This method facilitates the visualization of complex, high-dimensional datasets. The UMAP visualization in Fig. 4 elucidates the inherent structure and dispersion of the clusters, providing a visual confirmation of the distinct groupings identified within the multidimensional dataset.

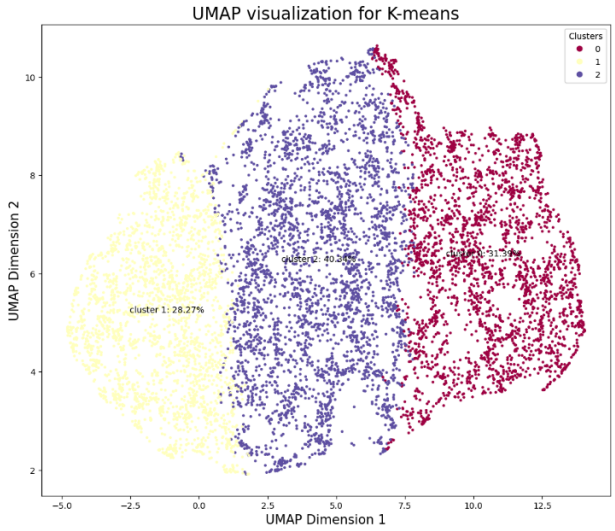

Fig. 4. UMAP visualization for K-means

Cluster 1, rendered in vibrant yellow, comprises 28.27% of the data points and is positioned primarily in the negative quadrant of UMAP Dimension 2, showing a clear delineation from the other clusters. This cluster's isolation could imply a set of topics or sentiments that are distinctly separate from the rest, potentially indicating a unique subset of experiences or opinions that stand apart from the common themes. Cluster 2, the largest group, is depicted in blue and represents 40.30% of the topics. It occupies a dense region in the positive quadrant of both UMAP dimensions, suggesting a significant convergence of shared topics within this cluster. The density and central positioning of Cluster 2 indicate that the experiences it represents are perhaps the most widespread among the reviews analyzed. Cluster 3, shown in red, accounts for 31.39% of the data and is distributed along the positive axis of UMAP Dimension 1 while extending across both positive and negative values of UMAP Dimension 2. The spread of this cluster suggests a variety of topics that share some overlap with the blue cluster yet also extend into unique territory, indicating a blend of familiar and distinctive themes.

The subtle overlaps between clusters, particularly at the boundaries, indicate shared experiences or topics that do not belong exclusively to one cluster. These overlaps demonstrate the complexity of the tourist experience, where certain aspects cannot be entirely separated but instead exist on a continuum. Although quantitative metrics are crucial for identifying optimal cluster formations, they are not comprehensive in assessing the thematic relevance or comprehensibility of the clusters. To augment the interpretative depth and afford a contextually nuanced viewpoint, we advocate employing word clouds in Fig. 5 to visualize the clusters. This qualitative method facilitates a more intuitive grasp of the underlying topics, proving particularly advantageous for stakeholders interpreting results in a business context.

The word cloud for Cluster 0 illustrates a rich tapestry of terms that tourists associate with their shopping experiences in Bangkok's top venues, as derived from TripAdvisor reviews. The most prominent words, such as "food," "shopping," "market," "cheap," "local," and "restaurant," suggest that visitors to Bangkok are particularly impressed by the intersection of affordable shopping with diverse culinary experiences. The significant presence of the word "food" alongside "shopping" and "market" indicates that culinary offerings are a substantial part of the retail experience in Bangkok. Tourists enjoy the convenience of accessing various food options while engaging in shopping activities, perhaps indicative of the street food culture embedded within the shopping areas. The recurrence of the word "cheap" in close association with "market" and "local" suggests that tourists value the affordability of products in these shopping venues. Such a pattern might reflect a penchant for bargain hunting and the purchase of local goods, pointing to an appreciation of traditional Thai markets over more westernized shopping malls. The word "restaurant" is also noteworthy, highlighting that formal dining experiences are a meaningful part of the shopping journey for many visitors. The presence of words like "nice," "enjoy," "recommend," and "clean" suggests that beyond price and variety, tourists also place high importance on the quality of the experience and the environment of the shopping venues. The terms "river," "boat," and "ferry" may indicate that the geographical features of Bangkok, such as its famous canals and rivers, play a role in the shopping experience, perhaps by providing scenic travel routes to and from shopping destinations or offering unique floating market experiences. Lastly, the words "walk," "visit," "busy," and "crowd" suggest that walking is a primary mode of exploration for tourists in these areas and that these venues are bustling hubs of activity. The observation of such vibrant activity could imply that despite the crowds, the vibrancy and energy of these places contribute positively to the overall shopping atmosphere.

According to TripAdvisor reviews, the word cloud for Cluster 1 reveals a constellation of terms that resonate with tourists' experiences at Bangkok's top shopping venues. The significant words such as "food," "mall," "shopping," "restaurant," and "clothes," suggest that for visitors, the retail experience is deeply intertwined with dining and a wide variety of merchandise options. The prominence of "mall" in conjunction with "food court" and "restaurant" indicates that.

According to TripAdvisor reviews, the word cloud for Cluster 1 reveals a constellation of terms that resonate with tourists' experiences at Bangkok's top shopping venues. The significant words such as "food," "mall," "shopping," "restaurant," and "clothes" suggest that for visitors, the retail experience is deeply intertwined with dining and a wide variety of merchandise options. The prominence of "mall" in conjunction with "food court" and "restaurant" indicates that large shopping centers, offering an amalgamation of food services and retail stores, are a central part of the shopping experience in Bangkok. These terms signify the importance of convenience and diversity in dining options as integral components of the retail environment. The frequent mention of "clothes," "shoe," and "bag" highlights the focus on fashion-related shopping, pointing to a strong interest in Bangkok as a destination for purchasing apparel and accessories. The adjacency of words such as "bargain," "cheap," and "expensive" reflects the wide price range that shoppers encounter, accommodating both budget-conscious consumers and those looking for premium products.

(a) Cluster 0

(b) Cluster 1

(c) Cluster 2

Fig. 5. K-means word cloud

Moreover, the words "tourist," "Thai," and "Bangkok" alongside "souvenir" and "item" underscore the city's appeal as a cultural shopping destination where visitors seek to purchase items that represent the local culture and craftsmanship. Including "siam" and "paragon," possibly referring to specific well-known shopping centers in Bangkok, indicates that certain venues have strong brand recognition among tourists. The word "central" might emphasize the centrality of these shopping places within the city's geography or the tourists' itineraries. The terms "walk," "enjoy," "experience," and "visit" suggest that shopping is not merely transactional for visitors but also a leisure activity that contributes to the overall enjoyment of their stay in Bangkok. The shopping venues appear to provide a multi-sensory, engaging, enjoyable experience.

The analysis of Cluster 2's word cloud reveals a nuanced perspective of the shopping experience in Thailand as shared by tourists on TripAdvisor. Dominant words such as "food," "shopping," "mall," "expensive," "local," and "restaurant" underscore the integral role of dining and varied price points in the retail experience. The significant size of the word "food" alongside "restaurant" and "food court" suggests that dining options are paramount to the shopping experience, with visitors seeking a seamless blend of culinary exploration within their retail excursions. Such prominence of dining-related terms may indicate that shopping venues are not only places of purchase but also social and gastronomic hubs. The juxtaposition of "expensive" with "bargain" and "cheap" implies a diverse economic landscape within these shopping venues, catering to a broad spectrum of financial expectations and shopping intents. The presence of "mall" indicates a preference for the convenience and variety offered by larger shopping centers. The recurrence of "local" alongside "Thai" and "Bangkok" denotes an appreciation for authentic local purchases, signaling that tourists are interested in items that reflect the cultural and regional distinctiveness of Thailand. Terms like "walk," "visit," "enjoy," and "experience" emphasize the experiential nature of shopping, suggesting that it is viewed as an enjoyable activity rather than a mere transaction. The idea that shopping is an experience to be savored is reinforced by words such as 'beautiful,' 'fun,' and 'clean,' which convey that shopping venues' overall ambiance and environment contribute substantially to their appeal. The words "stall," "market," and "night" also appear prominently, indicating the popularity of open-air markets and night markets, known for their vibrant atmospheres and unique shopping experiences. These markets are likely valued for their traditional shopping experience, juxtaposed with the more modern "mall" shopping. Additionally, the presence of transportation-related terms like "boat," "taxi," "BTS station," and "ferry" suggests that accessibility and the ease of transit to and from these venues are essential factors for visitors.

Across all clusters, the reoccurrence of words such as "food," "shopping," "Bangkok," and "restaurant" indicates a strong interconnection between eating and shopping as part of the tourist experience in Thailand's shopping venues. The analysis shows that while there is a significant overlap in subjects across the word clouds, each cluster maintains distinct themes of the shopping experiences as perceived by international tourists. Cluster 0 emphasizes the value-driven and culturally rich street market experience. Cluster 1 highlights the modern mall environment, where diversity in shopping, dining, and cultural activities is paramount. Cluster

2 indicates a preference for an all-encompassing retail experience that seamlessly integrates convenience with the diverse and accessible nature of Thailand's shopping hubs. These insights are valuable for stakeholders in Thailand's retail and tourism sectors, highlighting the multifaceted appeal that attracts tourists to these destinations.

G. Discussion

The comparative assessment of PPSO, PSO, and GA for fine-tuning ELM-AE highlights the delicate balance between minimizing reconstruction loss and optimizing computational efficiency. Results indicate that the optimal selection of hidden nodes (300 to 400) and the penalty coefficient (C) are crucial for balancing accuracy with training time. While this analysis provides insight into the relationship between hyperparameters and model performance, the applicability of these findings to diverse datasets and various contexts necessitates further exploration. Future research should focus on developing adaptive optimization techniques tailored to the specific needs of different datasets. The comparison between ELM-AE and traditional autoencoders reveals a trade-off between computational speed and depth of feature extraction. ELM-AE offers significant reductions in training time, enabling quicker feature extraction but potentially sacrificing the depth of semantic analysis seen in traditional autoencoders. This distinction underscores the need to choose the model that best suits specific application needs, considering the balance between speed and feature richness.

Moreover, the integrated evaluation using K-Means clustering and silhouette scores has highlighted the superior clustering performance of the optimized LDA-BERT model over traditional text representation methods, offering a detailed understanding of cluster coherence and separation. This study emphasizes the value of silhouette scores in measuring clustering quality and suggests enhancing clustering outcomes through advanced initialization techniques. The empirical analysis through K-means and Agglomerative clustering of the optimized LDA-BERT model demonstrates the robustness of the clustering process, with the elbow method identifying an optimal cluster count at $k = 3$. Both clustering methods showed comparable coherence, with K-means slightly outperforming in cluster separation, reinforcing the integrity of the clustering outcomes.

However, this study acknowledges limitations in the clustering analysis, such as the potential oversimplification of textual data's complex relationships by relying solely on the elbow method and silhouette scores for optimal cluster determination. Moreover, while UMAP visualization and word cloud analysis offer insightful thematic cluster delineations, they may not fully address the high-dimensional data's complexity, suggesting further exploration into capturing nuanced inter-cluster dynamics.

V. CONCLUSION AND FUTURE WORK

In this study, we conducted an in-depth investigation into topic modeling from textual corpora, employing a novel hybrid approach that combines the strengths of BERT and LDA for unsupervised clustering. The application of PPSO for optimizing ELM-AE for feature extraction and dimensionality reduction proved to be efficacious, achieving an optimal balance between computational efficiency and loss minimization. Utilizing a preprocessed dataset of TripAdvisor reviews, our hybrid model, assessed through silhouette and coherence metrics, demonstrated superior performance in extracting relevant topics from the data. By employing K-means clustering and utilizing the Elbow Method for determining the optimal number of clusters, our approach effectively synthesized the probabilistic topic assignments from LDA with the contextual embeddings from BERT, preserving semantic integrity and generating contextually rich topic information.

Our proposed unified BERT-LDA clustering framework has been designed and implemented to extract meaningful topics from textual corpora, with our experiments validating its efficacy in producing more coherent topic clusters. This approach, mainly applied to the analysis of Thailand's tourism reviews, has provided actionable insights for stakeholders, highlighting the multifaceted attractions of the region's shopping destinations and enhancing reputation management and branding through the analysis of thematic content for subsequent sentiment analysis. While this research offers valuable insights, it also recognizes certain limitations, such as the potential for bias in content created by users and the complexity of accurately deciphering the emotions and contexts expressed in the reviews.

In the future, exploring BERT variants and Deep Neural Networks (DNNs) for topic inference promises to extend the research's applicability beyond its current focus on English-language reviews. By integrating multilingual models, the framework can be adapted to analyze datasets across various linguistic contexts, increasing its global relevance. Moreover, the potential application of this methodology across different sectors, including health, education, and politics, offers exciting avenues for future research. Such adaptations could facilitate mining sector-specific topics, thereby broadening the impact and utility of our topic modeling framework in understanding and addressing the nuanced demands of diverse fields.

ACKNOWLEDGMENT

The authors would like to express their gratitude for the support provided by Khon Kaen University Grant for ASEAN-GMS Countries Year of 2022, College of Computing, Khon Kaen University, Thailand.

REFERENCES

[1] Z. Xiang and U. Gretzel, "Role of social media in Online Travel Information Search," *Tourism Management*, vol. 31, no. 2, pp. 179–188, 2010, doi: 10.1016/j.tourman.2009.02.016.

[2] S. Pike, L. P. Dam, and A. Beatson, "Social media gratifications in the context of international travel planning: The use of the repertory test method," *Acta turistica*, vol. 31, no. 2, pp. 153–178, 2019, doi: 10.22598/at/2019.31.2.153.

[3] R. A. Hamid *et al*., "How smart is e-tourism? A systematic review of Smart Tourism Recommendation System Applying Data Management," *Computer Science Review*, vol. 39, p. 100337, 2021, doi: 10.1016/j.cosrev.2020.100337.

[4] M. F. ALmasoodi, S. Rahman, M. Basendwah, and A. N. ALfarra, "Leveraging Digital Transformation to enhance quality tourism services in Babylon City, Iraq," *International Journal of Sustainable Development and Planning*, vol. 18, no. 10, pp. 3195–3211, 2023, doi: 10.18280/ijsdp.181020.

[5] J. Žižka, F. Dařena, and A. Svoboda, “Introduction to text mining with Machine Learning,” *Text Mining with Machine Learning*, pp. 1–12, Oct. 2019, doi: 10.1201/9780429469275-1.

[6] Q. Li *et al*., “A survey on text classification: From traditional to deep learning,” *ACM Transactions on Intelligent Systems and Technology*, vol. 13, no. 2, pp. 1–41, Apr. 2022, doi: 10.1145/3495162.

[7] Q. He, K. Chang, E.-P. Lim, and J. Zhang, "Bursty feature representation for clustering text streams," *Proceedings of the 2007 SIAM International Conference on Data Mining*, pp. 491-496, 2007, doi: 10.1137/1.9781611972771.50.

[8] E. C. Garrido-Merchan, R. Gozalo-Brizuela, and S. Gonzalez-Carvajal, “Comparing Bert against traditional machine learning models in text classification,” *Journal of Computational and Cognitive Engineering*, vol. 2, no. 4, pp. 352-356, Apr. 2023, doi: 10.47852/bonviewjcce3202838.

[9] A. S. Alammary, “Bert models for Arabic Text Classification: A systematic review,” *Applied Sciences*, vol. 12, no. 11, p. 5720, Jun. 2022, doi: 10.3390/app12115720.

[10] M. Mishra and J. Viradiya, "Survey of Sentence Embedding Methods," *International Journal of Applied Science and Computations*, vol. 6, no. 3, pp. 592-592, 2019, doi: 10.13140/RG.2.2.21861.45289.

[11] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *Journal of machine Learning research*, vol. 3, pp. 993-1022, Jan 2003.

[12] M. A. Ferraria, V. A. Ferraria, and L. N. de Castro, "An investigation into different text representations to train an artificial immune network for clustering texts," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 8, no. 3, p. 55, 2023, doi: 10.9781/ijimai.2023.08.006.

[13] Y. Tay, M. Dehghani, D. Bahri, and D. Metzler, “Efficient transformers: A survey,” *ACM Computing Surveys*, vol. 55, no. 6, pp. 1–28, Dec. 2022, doi: 10.1145/3530811.

[14] M. H. Ahmed, S. Tiun, N. Omar, and N. S. Sani, "Short text clustering algorithms, application and challenges: A survey," *Applied Sciences*, vol. 13, no. 1, p. 342, 2022, doi: 10.3390/app13010342.

[15] E. Keogh and A. Mueen, "Curse of dimensionality," *Encyclopedia of Machine Learning and Data Mining*, pp. 314–315, 2017, doi: 10.1007/978-1-4899-7687-1_192.

[16] R. Wang, J. Bian, F. Nie, and X. Li, "Unsupervised Discriminative Projection for Feature Selection," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 2, pp. 942-953, 1 Feb. 2022, doi: 10.1109/TKDE.2020.2983396.

[17] V. Raunak, V. Gupta, and F. Metze, “Effective dimensionality reduction for word embeddings,” *Proceedings of the 4th Workshop on Representation Learning for NLP*, pp. 235-243, 2019, doi: 10.18653/v1/w19-4328.

[18] I. T. Jolliffe and J. Cadima, “Principal component analysis: A review and recent developments,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202, Apr. 2016, doi: 10.1098/rsta.2015.0202.

[19] K. Pearson, “LIII. on lines and planes of closest fit to systems of points in space,” *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, no. 11, pp. 559–572, Nov. 1901, doi: 10.1080/14786440109462720.

[20] A. E. Maxwell, T. A. Warner, and F. Fang, “Implementation of machine-learning classification in Remote Sensing: An applied review,” *International Journal of Remote Sensing*, vol. 39, no. 9, pp. 2784–2817, Feb. 2018, doi: 10.1080/01431161.2018.1433343.

[21] W. Jia, M. Sun, J. Lian, and S. Hou, "Feature dimensionality reduction: a review," *Complex & Intelligent Systems*, vol. 8, 2022, doi: 10.1007/s40747-021-00637-x.

[22] W. Chen, X. Chen, and Y. Lin, "Homogeneous Ensemble Extreme Learning machine autoencoder with mutual representation learning and manifold regularization for medical datasets," *Applied Intelligence*, vol. 53, no. 12, pp. 15476–15495, 2022, doi: 10.1007/s10489-022-04284-8.

[23] K. Sun, J. Zhang, C. Zhang, and J. Hu, "Generalized extreme learning machine autoencoder and a new deep neural network," *Neurocomputing*, vol. 230, pp. 374–381, 2017, doi: 10.1016/j.neucom.2016.12.027.

[24] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with Neural Networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006, doi: 10.1126/science.1127647.

[25] J. Cai, S. Wang, and W. Guo, "Unsupervised embedded feature learning for deep clustering with stacked sparse auto-encoder," *Expert Systems with Applications*, vol. 186, p. 115729, 2021, doi: 10.1016/j.eswa.2021.115729.

[26] Y. Zhu *et al*., "Representation learning with deep sparse auto-encoder for multi-task learning," *Pattern Recognition*, vol. 129, p. 108742, 2022, doi: 10.1016/j.patcog.2022.108742.

[27] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion," *Journal of Machine Learning Research*, vol. 11, pp. 3371-3408, 2010.

[28] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006, doi: 10.1016/j.neucom.2005.12.126.

[29] T. Liu, C. K. L. Lekamalage, G.-B. Huang, and Z. Lin, "Extreme learning machine for joint embedding and clustering," *Neurocomputing*, vol. 277, pp. 78–88, 2018, doi: 10.1016/j.neucom.2017.01.115.

[30] H. Yongyong and Z. Xiaoqiang, "Sparse representation preserving embedding based on extreme learning machine for process monitoring," *Transactions of the Institute of Measurement and Control*, vol. 42, no. 10, pp. 1895–1907, 2020, doi: 10.1177/0142331219898937.

[31] L. Kasun, H. Zhou, G.-B. Huang, and C.-M. Vong, "Representational Learning with ELMs for Big Data," *IEEE Intelligent Systems*, vol. 28, pp. 31–34, 2013.

[32] C. M. Wong, C. M. Vong, P. K. Wong, and J. Cao, "Kernel-Based Multilayer Extreme Learning Machines for Representation Learning," in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 3, pp. 757-762, March 2018, doi: 10.1109/TNNLS.2016.2636834.

[33] M. Eshtay, H. Faris, and N. Obeid, "Metaheuristic-based extreme learning machines: a review of design formulations and applications," *International Journal of Machine Learning and Cybernetics*, vol. 10, 2019, doi: 10.1007/s13042-018-0833-6.

[34] M. Gholamghasemi, E. Akbari, M. Asadpoor, and M. Ghasemi, "A new solution to the non-convex economic load dispatch problems using phasor particle swarm optimization," *Applied Soft Computing*, vol. 79, 2019, doi: 10.1016/j.asoc.2019.03.038.

[35] J. Rashid, S. M. Shah, and A. Irtaza, “An efficient topic modeling approach for text mining and information retrieval through K-means clustering,” *Journal of Engineering & Technology*, vol. 39, no. 1, pp. 213–222, Jan. 2020, doi: 10.22581/muet1982.2001.20.

[36] M. Alhawarat and M. Hegazi, “Revisiting K-means and topic modeling, a comparison study to cluster Arabic documents,” *IEEE Access*, vol. 6, pp. 42740–42749, 2018, doi: 10.1109/access.2018.2852648.

[37] T. Joachims, "A probabilistic analysis of the Rocchio algorithm with TFIDF for text categorization," *Proc. 14th Int. Conf. Mach. Learn*., pp. 143-151, 1997.

[38] D. Yan, K. Li, S. Gu, and L. Yang, "Network-Based Bag-of-Words Model for Text Classification," in *IEEE Access*, vol. 8, pp. 82641-82652, 2020, doi: 10.1109/ACCESS.2020.2991074.

[39] E. M. Rinke, T. Dobbrick, C. Löb, C. Zirn, and H. Wessler, "Expert-informed topic models for document set Discovery," *Communication Methods and Measures*, vol. 16, no. 1, pp. 39–58, 2021, doi: 10.1080/19312458.2021.1920008.

[40] L. Xiang, "Application of an improved TF-IDF method in literary text classification," *Advances in Multimedia*, vol. 2022, pp. 1–10, 2022, doi: 10.1155/2022/9285324

[41] R. Koragoankar, V. Kulkarni, and D. Naik, "Search Engine Using NLP Text Processing Techniques to Extract Most Relevant Search Results," *2023 14th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, pp. 1-7, 2023, doi: 10.1109/ICCCNT56998.2023.10307392.

[42] R. C. Belwal, S. Rai, and A. Gupta, "Text summarization using topic-based vector space model and semantic measure," *Information Processing & Management*, vol. 58, no. 3, p. 102536, 2021, doi: 10.1016/j.ipm.2021.102536.

[43] B. P. Bhopale and A. Tiwari, "Leveraging Neural Network Phrase Embedding Model For Query Reformulation In Ad-Hoc Biomedical Information Retrieval," *Malaysian Journal of Computer Science*, vol. 34, no. 2, pp. 151–170, Apr. 2021, doi: 10.22452/mjcs.vol34no2.2.

[44] Y. Wu, X. Wang, W. Zhao, and X. Lv, "A novel topic clustering algorithm based on graph neural network for question topic diversity," *Information Sciences*, vol. 629, pp. 685–702, 2023, doi: 10.1016/j.ins.2023.02.018.

[45] M. W. Akram *et al*., "A novel deep auto-encoder based linguistics clustering model for social text," *ACM Transactions on Asian and Low-Resource Language Information Processing*, 2022, doi: 10.1145/3527838.

[46] Y. Guo, R. Fei, K. Zhang, Y. Tang, and B. Hu, "Developing a Clustering Structure with Consideration of Cross-Domain Text Classification based on Deep Sparse Auto-encoder," *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2477-2483, 2020, doi: 10.1109/BIBM49941.2020.9313537.

[47] C. Sun *et al*., "A Deep Learning Approach With Deep Contextualized Word Representations for Chemical–Protein Interaction Extraction From Biomedical Literature," in *IEEE Access*, vol. 7, pp. 151034-151046, 2019, doi: 10.1109/ACCESS.2019.2948155.

[48] S. A. Curiskis, B. Drake, T. R. Osborn, and P. J. Kennedy, "An evaluation of document clustering and topic modelling in two online social networks: Twitter and reddit," *Information Processing & Management*, vol. 57, no. 2, p. 102034, 2020, doi: 10.1016/j.ipm.2019.04.002.

[49] Y. Liu *et al*., "RoBERTa: A robustly optimized BERT pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.

[50] T. B. Brown *et al*., "Language models are few-shot learners," in *Proc. 34th Int. Conf. Neural Inf. Process. Syst. (NIPS'20)*, pp. 1–25, 2020.

[51] C. Raffel *et al*., "Exploring the limits of transfer learning with a unified text-to-text transformer," *Journal of Machine Learning Research*, vol. 21, no. 1, pp. 5485–5551, 2020.

[52] P. He, X. Liu, J. Gao, and W. Chen, "DeBERTa: Decoding-enhanced BERT with Disentangled Attention," *arXiv preprint arXiv:2006.03654*, 2020, doi: 10.48550/arXiv.2006.03654.

[53] Z. Dai and J. Callan, "Deeper text understanding for IR with contextual neural language modeling," *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 985-988, 2019, doi: 10.1145/3331184.3331303.

[54] X. Ma *et al*., "B-PROP: bootstrapped pre-training with representative words prediction for ad-hoc retrieval," in *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 1513-1522, 2021, doi: 10.1145/3404835.3462869.

[55] Q. Liu *et al*., "Sgat: A self-supervised graph attention network for biomedical relation extraction," *2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2021, doi: 10.1109/bibm52615.2021.9669699.

[56] L. Huang, P. Shi, H. Zhu, and T. Chen, "Early detection of emergency events from social media: A new text clustering approach," *Natural Hazards*, vol. 111, no. 1, pp. 851–875, 2022. doi:10.1007/s11069-021-05081-1.

[57] Y. Zhou, C. Li, S. He, X. Wang, and Y. Qiu, "Pre-trained contextualized representation for Chinese conversation topic classification," *2019 IEEE International Conference on Intelligence and Security Informatics (ISI)*, 2019, doi: 10.1109/isi.2019.8823172.

[58] B. Chen, Z. Xu, X. Wang, L. Xu, and W. Zhang, "Capsule Network-based text sentiment classification," *IFAC-PapersOnLine*, vol. 53, no. 5, pp. 698–703, 2020, doi: 10.1016/j.ifacol.2021.04.160.

[59] D. Yamunathangam, C. B. Priya, G. Shobana, and L. Latha, "An Overview of Topic Representation and Topic Modelling Methods for Short Texts and Long Corpus," *2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)*, pp. 1-6, 2021, doi: 10.1109/ICAECA52838.2021.9675579.

[60] B. Chen, L. Fan, and X. Fu, "Sentiment Classification of Tourism Based on Rules and LDA Topic Model," *2019 International Conference on Electronic Engineering and Informatics (EEI)*, pp. 471-475, 2019, doi: 10.1109/EEI48997.2019.00108.

[61] A. -G. Văduva, M. Munteanu, S. -V. Oprea, A. Bâra, and A. -M. Niculae, "Understanding Climate Change and Air Quality Over the Last Decade: Evidence From News and Weather Data Processing," in *IEEE Access*, vol. 11, pp. 144631-144648, 2023, doi: 10.1109/ACCESS.2023.3345466.

[62] Q. Xie, X. Zhang, Y. Ding, and M. Song, "Monolingual and multilingual topic analysis using LDA and Bert Embeddings," *Journal of Informetrics*, vol. 14, no. 3, p. 101055, 2020, doi: 10.1016/j.joi.2020.101055.

[63] E. Atagun, B. Hartoka, and A. Albayrak, "Topic modeling using LDA and Bert Techniques: Teknofest example," *2021 6th International Conference on Computer Science and Engineering (UBMK)*, 2021, doi: 10.1109/ubmk52708.2021.9558988.

[64] I. Ali and M. A. Naeem, "Identifying and Profiling User Interest over time using Social Data," *2022 24th International Multitopic Conference (INMIC)*, pp. 1-6, 2022, doi: 10.1109/INMIC56986.2022.9972955.

[65] K. Kukushkin, Y. Ryabov, and A. Borovkov, "Digital Twins: A systematic literature review based on data analysis and Topic modeling," *Data*, vol. 7, no. 12, p. 173, 2022, doi: 10.3390/data7120173.

[66] R. S. Nambiar and D. Gupta, "Dedicated Farm-Haystack Question Answering System for Pregnant Women and Neonates Using Corona Virus Literature," *2022 12th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, pp. 222-227, 2022, doi: 10.1109/Confluence52989.2022.9734125.

[67] M. Khder, "Web scraping or web crawling: State of Art, Techniques, approaches and application," *International Journal of Advances in Soft Computing and its Applications*, vol. 13, no. 3, pp. 145–168, 2021, doi: 10.15849/ijasca.211128.11.

[68] D. M. Blei, "Probabilistic topic models," *Communications of the ACM*, vol. 55, no. 4, pp. 77–84, 2012, doi: 10.1145/2133806.2133826.

[69] U. Chauhan and A. Shah, "Topic modeling using latent Dirichlet allocation," *ACM Computing Surveys*, vol. 54, no. 7, pp. 1–35, 2021, doi: 10.1145/3462478.

[70] C. Li *et al*., "Mining Dynamics of research topics based on the combined LDA and WordNet," *IEEE Access*, vol. 7, pp. 6386–6399, 2019, doi: 10.1109/access.2018.2887314.

[71] D. Mimno, H. Wallach, E. Talley, M. Leenders, and A. Mccallum, "Optimizing Semantic Coherence in Topic Models," in *EMNLP 2011 - Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pp. 262–272, 2011.

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

[73] A. Vaswani *et al*., "Attention is all you need," in *Advances in Neural Information Processing Systems*, pp. 6000-6010, 2017.

[74] A. Karimi, L. Rossi, and A. Prati, "Adversarial Training for Aspect-Based Sentiment Analysis with BERT," *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 8797-8803, 2021, doi: 10.1109/ICPR48806.2021.9412167.

[75] S. Lacoste-Julien, F. Sha, and M. Jordan, "DiscLDA: Discriminative Learning for Dimensionality Reduction and Classification," in *Advances in Neural Information Processing Systems*, pp. 897-904, 2008.

[76] J. A. Lossio-Ventura, J. Morzan, H. Alatrista-Salas, T. Hernandez-Boussard, and J. Bian, "Clustering and topic modeling over tweets: A comparison over a health dataset," *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 1544-1547, 2019, doi: 10.1109/BIBM47256.2019.8983167.

[77] G. Tang, X. Chen, N. Li, and J. Cui, "Research on the evolution of journal topic mining based on the bert-LDA model," *SHS Web of Conferences*, vol. 152, p. 03012, 2023, doi: 10.1051/shsconf/202315203012.

[78] B. Deng, X. Zhang, W. Gong, and D. Shang, "An Overview of Extreme Learning Machine," *2019 4th International Conference on Control, Robotics and Cybernetics (CRC)*, pp. 189-195, 2019, doi: 10.1109/CRC.2019.00046.

[79] O. Ali, Q. Abbas, K. Mahmood, E. Thompson, J. Arambarri, and I. Ashraf, "Competitive Coevolution-Based Improved Phasor Particle Swarm Optimization Algorithm for Solving Continuous Problems," *Mathematics*, vol. 11, p. 4406, 2023, doi: 10.3390/math11214406.

[80] M. Ghasemi *et al*., "Phasor particle swarm optimization: A simple and efficient variant of PSO," *Soft Computing*, vol. 23, no. 19, pp. 9701–9718, 2018, doi: 10.1007/s00500-018-3536-8.

[81] M. Gajić, M. Jevtic, J. Radosavljević, S. Arsic, and D. Klimenta, "Phasor Particle Swarm Optimization for Solving Problem of Pricing in Electricity Market," in *International Scientific Conference "UNITECH*, vol. 1, p. 257, 2021.

[82] D. El Bourakadi, A. Yahyaouy, and J. Boumhidi, "Improved extreme learning machine with AutoEncoder and particle swarm optimization for short-term wind power prediction," *Neural Computing and Applications*, vol. 34, pp. 1–17, 2022, doi: 10.1007/s00521-021-06619-x.

[83] G. H. de Rosa, J. R. Brega, and J. P. Papa, "How optimizing perplexity can affect the dimensionality reduction on word embeddings visualization?" *SN Applied Sciences*, vol. 1, no. 12, Nov. 2019, doi: 10.1007/s42452-019-1689-4.

[84] Z. Tian, Y. Ren, and G. Wang, "Short-term wind speed prediction based on improved PSO algorithm optimized EM-ELM," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, vol. 41, pp. 26–46, 2019.

[85] Y. Wang, D. Wang, and Y. Tang, "Clustered Hybrid Wind Power Prediction Model Based on ARMA, PSO-SVM, and Clustering Methods," in *IEEE Access*, vol. 8, pp. 17071-17079, 2020, doi: 10.1109/ACCESS.2020.2968390.

[86] M. Srinivas and L. M. Patnaik, "Genetic algorithms: a survey," in *Computer*, vol. 27, no. 6, pp. 17-26, 1994, doi: 10.1109/2.294849.

[87] M. Ahmed, R. Seraj, and S. M. Islam, "The K-Means Algorithm: A comprehensive survey and performance evaluation," *Electronics*, vol. 9, no. 8, p. 1295, 2020, doi:10.3390/electronics9081295.

[88] K. Sethia, M. Saxena, M. Goyal, and R. K. Yadav, "Framework for Topic Modeling using BERT, LDA and K-Means," *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, pp. 2204-2208, 2022, doi: 10.1109/ICACITE53722.2022.9823442.

[89] L. George and P. Sumathy, "An Integrated Clustering and BERT Framework for Improved Topic Modeling," pp. 1-9, 2022, doi: 10.21203/rs.3.rs-1986180/v1.

[90] T. Dinh, T. Fujinami, and V.-N. Huynh, "Estimating the Optimal Number of Clusters in Categorical Data Clustering by Silhouette Coefficient," in *Proceedings of the International Conference on Advanced Technologies for Communications*, pp. 1-1, 2019, doi: 10.1007/978-981-15-1209-4_1.

[91] K. R. Shahapure and C. Nicholas, "Cluster Quality Analysis Using Silhouette Score," *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, pp. 747-748, 2020, doi: 10.1109/DSAA49011.2020.00096.

[92] A. Naghizadeh and D. N. Metaxas, "Condensed silhouette: An optimized filtering process for cluster selection in K-means," in *Procedia Computer Science*, vol. 176, pp. 205-214, 2020.

[93] I. Bouabdallaoui, F. Guerouate, and M. Sbihi, "Combination of genetic algorithms and k-means for a hybrid topic modeling: Tourism use case," *Evolutionary Intelligence*, pp. 1-17, 2023, doi: 10.1007/s12065-023-00863-x.

[94] D. Marutho, S. H. Handaka, E. Wijaya, and Muljono, "The Determination of Cluster Number at k-Mean Using Elbow Method and Purity Evaluation on Headline News," *2018 International Seminar on Application for Technology of Information and Communication*, pp. 533-538, 2018, doi: 10.1109/ISEMANTIC.2018.8549751.

[95] L. Liu *et al*., "Fast identification of urban sprawl based on K-means clustering with population density and local spatial entropy," *Sustainability*, vol. 10, no. 8, p. 2683, 2018, doi: 10.3390/su10082683.

[96] A. H. Marani and E. P. Baumer, "A review of stability in topic modeling: Metrics for assessing and techniques for improving stability," *ACM Computing Surveys*, vol. 56, no. 5, pp. 1–32, 2023, doi: 10.1145/3623269.

[97] A. El-Hamdouchi, "Comparison of hierarchic agglomerative clustering methods for document retrieval," *The Computer Journal*, vol. 32, no. 3, pp. 220–227, 1989, doi: 10.1093/comjnl/32.3.220.

[98] B. Ghojogh, M. Crowley, F. Karray, and A. Ghodsi, "Uniform manifold approximation and projection (UMAP)," *Elements of Dimensionality Reduction and Manifold Learning*, pp. 479–497, 2023, doi: 10.1007/978-3-031-10602-6_17.