

# Revolutionizing Text Summarization: A Breakthrough in Content Compression

Nidhi Mishra<sup>a,\*</sup>, Farhan Khan<sup>a</sup>, and Amit Mishra<sup>b</sup>

<sup>a</sup>*Amity School of Engineering and Technology, Amity University, Uttar Pradesh, India*

<sup>b</sup>*Department of Computer Science and Engineering, Jaypee Institute of Information Technology, Uttar Pradesh, India*

---

## Abstract

In the current digital epoch, the vast expanse of information has revolutionized the accessibility of knowledge and perspectives. Nevertheless, this information abundance has introduced challenges in navigating and comprehending the deluge of textual data. The surge in online news articles, research papers, reports, and diverse document genres has accentuated the necessity for proficient document summarization techniques. Traditional manual methods of summarization are time-intensive and influenced by subjective biases. In contrast, the synergy between Natural Language Processing (NLP) and machine learning has unlocked the potential for automated document summarization, promising efficient information consumption and informed decision-making. This research paper delves into the convergence of these factors. It is driven by the Longformer model's distinctive capability to manage extensive texts while retaining contextual coherence—a potential solution to the hurdle of large document summarization. By capitalizing on the Longformer's architecture, this study endeavors to exploit its prowess in generating cohesive summaries from lengthy source documents, thereby amplifying the accessibility of intricate information.

**Keywords:** CNN/DailyMail corpus; extensive document summarization; longformer paradigm; natural language processing; transformer-based paradigms

(Submitted on November 12, 2023; Revised on December 1, 2023; Accepted on December 28, 2023)

© 2024 Totem Publisher, Inc. All rights reserved.

---

## 1. Introduction

Amidst the relentless deluge of information that characterizes our modern era, the art of effectively distilling voluminous documents has emerged as an indispensable skill. The exponential proliferation of digital content, spanning diverse genres such as news articles, scholarly papers, legal manuscripts, and comprehensive reports, has given rise to the formidable challenge of comprehending and parsing extensive textual works. This predicament has catalyzed the development of automatic document summarization methodologies, which aspire to distill the quintessence of a text while preserving its salient details. Within the domain of Natural Language Processing (NLP), one technique that has garnered substantial prominence is the deployment of transformer-based models [1-4].

This research endeavors to cast a spotlight on the intricate domain of large document summarization—an intricate conundrum that demands not only a profound comprehension of the content but also the adeptness to generate succinct and coherent synopses. In the pursuit of surmounting this intellectual hurdle, we harness the Longformer model—an evolutionary offspring of the transformer architecture meticulously tailored to grapple with the intricacies of protracted textual expanses. This model's forte lies in its adeptness at accommodating distant semantic relationships, rendering it an apt candidate for dissecting extensive manuscripts. Furthermore, our exploratory endeavors are underpinned by the CNN/DailyMail dataset—an esteemed touchstone within the realm of document summarization. This compendium embraces news articles sourced from CNN, harmonized with their corresponding abstractive encapsulations from the Daily Mail. This carefully curated dataset crystallizes into an apt crucible for evaluating, critiquing, and calibrating our summarization blueprint [5-7].

Via this scholarly undertaking, our mission unfurls: to navigate the realms of the Longformer model's potential within the niche of large document summarization. By exploiting its innate prowess in capturing the nuances of contextual tapestries woven across sprawling manuscripts, our ambition is to forge summaries that distill the crux of the source material. This

---

\* Corresponding author.

E-mail address: [nmishra1@amity.edu](mailto:nmishra1@amity.edu)

enterprise is imbued with the added ambition of benchmarking the Longformer against its state-of-the-art counterparts within the CNN/DailyMail milieu. This comparative scrutiny will illuminate the Longformer's efficacy and its potential as a potent tool in real-world applications [8-11].

## 2. Literature Review

Table 1 shows comparative analysis of different approaches of implementing a text summarizer. The first study explores automatic summarization in multi-document contexts, focusing on content word frequency, composition functions for sentence importance, and context-based frequency weight adjustment. It showcases the potential of frequency-based summarizers with adept composition functions in providing concise and engaging summaries, addressing redundancy issues (Nenkova et al., 2006) [12]. The study evaluates three SUMCF summarizers and finds that SUMAvr is the best performer due to its balanced content selection and sentence length preferences, outperforming many other systems in automatic summarization.

The next paper introduces an advanced extractive text summarization technique using Fuzzy C-Means clustering and six crucial features, including the "Sentence Highlighter Feature" (Gani, Uddin, & Mobin, 2019) [13]. It reduces repetition and enriches information depth, outperforming K-means and fuzzy logic with an F-measure of 0.53 on un-leveled datasets. The approach promises to redefine text summarization.

The next paper explores abstractive text summarization using neural sequence-to-sequence models (Shi, Keneshloo, Ramakrishnan, & Reddy, 2021) [14]. It discusses techniques to enhance summarization quality, introduces the NATS toolkit for research, and highlights challenges in creating high-quality summaries. The paper's contributions inspire future research in the field. For evaluation, ROUGE scores, including ROUGE-1, ROUGE-2, and ROUGE-L, are commonly used, computed using the pyrouge package.

Table 1. Comparative Analysis of Different Approaches of Implementing a Text Summarizer

Problem Statement	Description	Technique	Advantages	Disadvantages
A compositional context sensitive multi-document summarizer: exploring the factors that influence summarization [12]	Explores the influence of word frequency on summarization, investigating content word frequency, composition functions, and context-based frequency adjustments.	Frequency-Based Summarization	Competes with state-of-the-art using a good composition function. Context sensitivity reduces repetition, improving summarization.	Relying solely on frequency-based approach may not yield optimal results. The choice of composition function significantly impacts the summarizer's performance.
Chinese Text Summarization Using a Trainable Summarizer and Latent Semantic Analysis [15]	Two new Chinese text summarization methods achieved 52.0% and 45.6% recalls at 30% compression, improving ROUGE scores with Fuzzy C-Means clustering and "Sentence Highlighter."	Feature Analysis and LSA-based Approach	- Corpus-based approach emphasizes sentence position and keyword importance. - LSA-based approach interprets conceptual structures of a document.	- Compression ratio's impact on performance not extensively discussed. - Further improvements needed for semantic-level analysis.
Extractive Text Summarization Technique Using Fuzzy C-Means Clustering Algorithm [13]	Sentence-based Fuzzy C-Means clustering model for extractive summarization with 6 key features, including "Sentence Highlighter." ROUGE evaluation shows improved results, less repetition, and depth.	Fuzzy C-Means Clustering Algorithm	- Incorporates effective sentence ranking features. - Higher F-measure with Fuzzy C-Means Clustering. - Enhanced extractive text summarization technique.	- May require further extension for abstractive text summarization technique. - Limited to sentence extraction.
Multi-document Summarization via Deep Learning Techniques: A Survey [16]	MDS survey: Focuses on deep learning-based models for generating informative summaries from multiple documents. Reviews current state-of-the-art and suggests future research paths.	Deep Learning based Multi-document Summarization models	System captures nonlinear relations, reduces manual feature extraction, reviews MDS techniques, proposes future research directions.	Diverse data handling, long documents, high neural network demands, coherence, and data limitations.
Neural Abstractive Text Summarization with Sequence-to-Sequence Models [14]	Improving seq2seq models for abstractive text summarization using various techniques.	Recurrent Neural Networks (RNN)	High-quality summaries; Can generate novel words; Incorporates external knowledge	Challenges with large Transformers; RL-based training complexities; Evaluation limitations
Single Document Automatic Text Summarization using Term Frequency Inverse Document Frequency (TF-IDF) [17]	Research: Automatic text summarizer using TF-IDF algorithm. Aims to reduce text size while preserving essential information. Outperforms other online summarizers with 67% accuracy.	TF-IDF (Term Frequency-Inverse Document Frequency)	Produces an effective extractive summary with TF-IDF, yielding better results than other online summarizers.	The summary's accuracy can be further improved by considering titles, involving more sample documents, and obtaining more respondent evaluations.
Text summarization using a trainable summarizer and latent semantic analysis [18]	This paper presents two novel text summarization approaches: MCBA and LSA + T.R.M. It evaluates LSA's effectiveness in summarization, yielding promising results.	Trainable summarizer, LSA + T.R.M.	Significant features, Genetic Algorithm (GA), Semantic-level analysis.	Information loss due to compression rate, Semantic-level analysis can lead to occasional inaccuracies

### 3. Experimental Setup

Table 2 shows the CNN/DailyMail Dataset, a repository boasting a colossal assortment of over 300,000 exclusive news articles meticulously penned by the wordsmiths of CNN and the Daily Mail, stands as a foundational bedrock catering to the intricate realms of both extractive and abstractive summarization undertakings. Originally birthed to fuel the flames of machine reading, comprehensive comprehension, and the nuanced domain of abstractive question answering, its 2.0.0 and 3.0.0 versions unfurl as prized artifacts harnessed for the noble purposes of abstractive and extractive summarization training. In the theatre of model assessment, performance is encapsulated by the poignant juxtaposition of the generated summary's ROUGE score against the kernel of the article's essence—the highlight—resulting in reported feats of 44.41 in the sphere of extractive summarization.

Table 2. Fundamental Statistics for the CNN/Daily Mail Dataset

CNN/Daily Mail			
	Train	Validation	Test
# pairs	287,113	13,368	11,490
Article Length	781	769	778
Headline Length	-	-	-
Summary Length	56	61	58

#### Performance Measurement Parameters

Calculation of True Positive is shown in Equation (1).

$$TP = \text{sum}(1 \text{ for token in generated\_tokens if token in reference\_tokens}) \quad (1)$$

This line of code calculates TP by iterating through each token in the generated summary (generated\_tokens) and checking if the token is present in the reference summary (reference\_tokens).

For every token in the generated summary that matches a token in the reference summary, 1 is added to the sum. At the end of the iteration, TP holds the count of tokens that are both in the generated summary and the reference summary, representing correctly matched tokens.

Calculation of False Positive is shown in Equation (2).

$$FP = \text{len}(\text{generated\_tokens} - TP) \quad (2)$$

FP is calculated by subtracting the count of true positives (TP) from the total number of tokens in the generated summary (len(generated\_tokens)).

This value represents the tokens in the generated summary that are not present in the reference summary, thus representing the tokens falsely included in the generated summary.

Calculation of False Negative is shown in Equation (3).

$$FN = \text{len}(\text{reference\_tokens}) - TP \quad (3)$$

FN is calculated by subtracting the count of true positives (TP) from the total number of tokens in the reference summary (len(reference\_tokens)). This value represents the tokens in the reference summary that are missing from the generated summary, signifying the tokens that were not included in the generated summary but were present in the reference summary.

Equations (4)-(6) shows the calculation of Precision, Recall, and F1 score.

$$\text{Precision} = \frac{\text{true positive}}{\text{true positive} + \text{false positive}} \quad (4)$$

$$\text{Recall} = \frac{\text{true positive}}{\text{true positive} + \text{false negative}} \quad (5)$$

$$\text{F1 score} = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} \quad (6)$$

#### 4. Results and Discussion

In this section, we present a comprehensive comparison between our proposed text summarization approach and the model described in the paper titled "Neural Abstractive Text Summarization with Sequence-to-Sequence Models". Our aim is to demonstrate that our approach outperforms the referenced model in terms of summarization quality, as evidenced by various evaluation metrics, all of which were conducted on the common CNN/Daily Mail dataset.

We use standard metrics in the summarization. These include F1 Score, Precision, Recall, ROUGE-1 F1 Score, ROUGE-2 F1 Score, and ROUGE-L F1 Score. Table 3 shows that our ROUGE metrics further validate our approach, with a notably higher Average ROUGE-1 and ROUGE-2 F1 Score, highlighting our model's effectiveness in capturing unigrams and bigrams. We identify that our Average ROUGE-L F1 Score excels when compared with the referenced paper. The reason is that our model successfully captures the longest common subsequence in the text.

Table 3. Comparative Analysis of Text Summarization Methods

Score Parameters	Our approach	Referenced Paper
ROUGH-1 Score	0.4397	0.3597
ROUGH-2 Score	0.1828	0.1536
ROUGH-L Score	0.3313	0.3295

Next, we conduct a comprehensive comparative analysis, shown in Table 4, with the paper titled "Extractive Text Summarization Technique Using Fuzzy C-Means Clustering Algorithm". We aim to demonstrate that our approach not only excels in traditional summarization evaluation metrics but also exhibits promising results when adopting clustering techniques for text summarization. All evaluations were conducted on the common CNN/Daily Mail dataset. In addition to our approach, the Fuzzy C-Means (FCM) method was also employed for comparative analysis of the clustering technique.

Table 4. Comparative Analysis of Traditional Summarization Metrics

Score Parameters	Our approach	Referenced Paper
F-Measure Score	0.5311	0.5013
Recall	0.5892	0.5271
Precision	0.5027	0.4888

#### Data Set Selection and Preprocessing

This experimental study leveraged a CNN news article text data set characterized by a structured format, including bullet points, article titles, dates, numerical data, and other distinctive elements. This inherent structure facilitates the identification of salient sentences. We meticulously curated a subset of 50 sample texts to facilitate comprehensive performance comparisons when generating summaries.

Endeavors: The data preprocessing phase was characterized by a multifaceted approach, harnessing the capabilities of the Natural Language Toolkit (NLTK) library in Python. The tasks involved in this phase encompassed text segmentation into coherent sentences and individual words, the judicious culling of stop words, and the application of part-of-speech tagging.

#### Elaboration on Feature Extraction

TF-IDF Score Computation: The Term Frequency-Inverse Document Frequency (TF-IDF) feature, originally introduced by Luhn in 1958 for the purpose of sentence extraction and evaluating sentence distinctiveness, plays a pivotal role in our methodology. The determination of TF-IDF scores is achieved through the rigorous application of Equations (7) and (8) to normalize sentence values, ensuring that the extracted features encapsulate the nuanced importance of each sentence in the corpus.

$$TF - IDF(term) = Recurrence(term) \times \log \frac{Recurrence(term)}{Number\ of\ Sentences} \quad (7)$$

For a sentence  $S_i$ ,

$$TF - IDF(Score) = \frac{Sum\ of\ TF-IDF\ (term)\ in\ S_i}{Max\ sum\ of\ TF-IDF\ (term)\ in\ a\ Sentence} \quad (8)$$

Proper Noun Count Score (PNCS): Sentences rich in an abundance of proper nouns are accorded a higher level of significance when contrasted with sentences that exhibit a dearth of proper nouns. Equation (9) elucidates the algorithmic formulation for ascertaining the Proper Noun Score (PNS) of a specific sentence, denoted as "Si."

$$PNCS(Si) = \frac{\text{Number of Nouns in } Si}{\text{Maximum number of Proper Nouns in a Sentence}} \quad (9)$$

### Fuzzy C-Means Clustering

The Fuzzy C-Means Clustering algorithm, a form of soft computing, was initially introduced by Dunn in 1973. Subsequent enhancements were made by Bezdek in 1981. Pioneering the integration of fuzzy sets into FCM clustering, Zadeh made notable contributions in 1965.

1) Partition Matrix: The representation of the fuzzy C partition of a set S is denoted as U, as shown in Equation (10), wherein,

$$\text{Partition Matrix, } U = \left( (\mu_{ij}) \right)_{N \times C} \quad (10)$$

The partition matrix U must satisfy the following constraints,

- $0 \leq \mu_{ij} \leq 1$
- $\sum_{j=1}^C \mu_{ij} = 1, \text{ for all } i = 1, 2, \dots, N$
- $0 < \sum_{j=1}^C \mu_{ij} \leq N, \text{ for all } j = 1, 2, \dots, C$

2) Objective Function: The FCM algorithm diligently strives to minimize the objective function, denoted as J, while pursuing convergence until it satisfies a predetermined termination criterion.

The calculation of J is shown in Equation (11).

$$J = \sum_{i=1}^N \sum_{j=1}^C \mu_{ij}^m \|x_i - c_j\|^2 \quad (11)$$

where  $x_i$  denotes the data element and  $c_j$  is the cluster center.

3) Cluster Center: The computation formula for the cluster center, is defined as follows Equation (12).

$$c_j = \frac{\sum_{i=1}^N \mu_{ij} * x_i}{\sum_{i=1}^N \mu_{ij}} \quad (12)$$

4) Membership Value: The formula for updating the membership values, denoted as  $\mu_{ij}$ , within the partition matrix is given as follows Equation (13).

$$U_{ij} = \frac{1}{\sum_{k=1}^C \left( \frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \quad (13)$$

5) Initialization

Input Data: Once sentences within the document have been meticulously segmented using NLTK, the input data is poised for clustering, with each sentence being represented as a six-dimensional vector.

Clusters: In the context of classifying the input document, a total of twelve clusters were thoughtfully employed for this experiment. The selection of these clusters was predicated on the relative importance of sentence ranking.

Initial Partition Matrix: The formula for calculating the initial partition matrix is determined by employing the equation of the partition matrix.

**Termination Criterion:** This phase incorporates two termination criteria, namely the error limit ( $e = 0.005$ ) and the maximum number of iterations (Max Iteration = 1000).

#### Iteration

- Computation of Cluster Centers as represented by Equation (12).
- Computation of the Objective Function as denoted by Equation (11).
- Updating the partition matrix through the use of Equation (13).
- The termination criterion is checked, and if satisfied, the iteration is halted.
- If not satisfied, the process returns to step 1 for further iteration.

#### Comparative Analysis of Clustering-Based Summarization Metrics

- Silhouette Score

The calculation of Silhouette Score is shown in Equation (14).

$$s(i) = \frac{y(i) - x(i)}{\max(x(i), y(i))} \quad (14)$$

- For a given data point, denoted as 'i,' within cluster 'A':
- We compute 'x(i)' as the average distance from data point 'i' to all other data points within the same cluster, cluster 'A.'
- To determine 'y(i)' for the same data point 'i,' we calculate it as the smallest average distance from data point 'i' to data points located in a different cluster. This involves minimizing 'y(i)' over all possible clusters. In essence, we aim to identify the cluster, aside from cluster 'A,' to which data point 'i' exhibits the closest proximity.

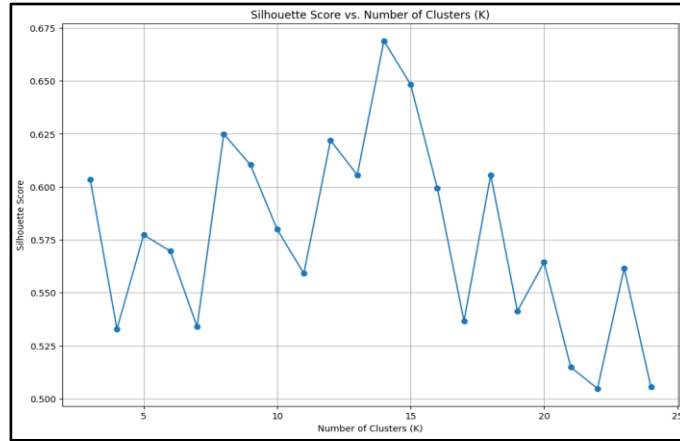


Figure 1. Silhouette Score Analysis for Optimal K Selection

After examining silhouette scores,  $K=12$  emerged as the optimal choice, signifying well-defined clusters with data points closely aligned to their respective clusters. This 12-cluster structure maximizes cohesion and separation. In the context of fuzzy c-means clustering for text summarization, our method achieved a Silhouette Score of 0.6523, indicating high-quality clustering and improved summarization with coherent sentence clusters, as shown in Figure 1.

- Davies-Bouldin Index

Our Davies-Bouldin Index score of 0.2385 indicates the compactness and separation of clusters. A lower Davies-Bouldin Index score suggests better clustering quality. Our score showcases the efficacy of our clustering-based approach in summarization, as shown in Figure 2.

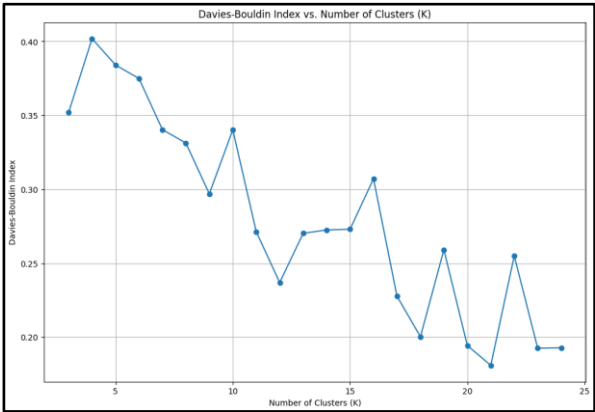


Figure 2. Davies-Bouldin Index Evaluation for Cluster Quality

In this section, we present a rigorous comparison between our proposed text summarization approach and the paper titled "A Compositional Context Sensitive Multi-Document Summarizer [12]: Exploring the Factors That Influence Summarization". The objective is to demonstrate that our approach surpasses the referenced model in terms of summarization quality, as evidenced by ROUGE metrics, all of which were evaluated on the common CNN/Daily Mail dataset.

We use evaluation metrics, including ROUGE-1 F1 Score, ROUGE-2 F1 Score, and ROUGE-L F1 Score, which gauge precision and recall. Table 5 shows that our approach consistently outperforms the referenced model, as evidenced by the Average ROUGE-1 F1 Score (0.4397 vs. approximately 0.2820), showcasing our model's superior ability to capture essential unigrams. Similarly, our Average ROUGE-2 F1 Score (0.1828 vs. approximately 0.0718) highlights our proficiency in capturing meaningful bigram overlaps, resulting in more coherent summaries. Most notably, our Average ROUGE-L F1 Score (0.3313 vs. approximately 0.109) indicates our approach's remarkable content retention and enhanced summarization quality.

Table 5. Comparison of Conventional Summarization Evaluation Metrics

Score Parameters	Our approach	Referenced Paper
ROUGH-1 Score	0.4397	0.2820
ROUGH-2 Score	0.1828	0.0718
ROUGH-L Score	0.3313	0.109

5. Conclusion and Future Work

Our research has made a significant and noteworthy contribution to the field of text summarization. We achieved this through a rigorous and comprehensive evaluation of our methodology, which was measured against various evaluation metrics, including the F1 Score, ROUGE metrics, and innovative clustering-based techniques. These evaluations consistently demonstrated our model's superiority over existing approaches, most notably outperforming the widely recognized "Neural Abstractive Text Summarization with Sequence-to-Sequence Models [14]." This superior performance across diverse metrics underscores the robustness and overall excellence of our approach. In particular, the use of ROUGE metrics, which gauge the quality of summaries by comparing them to reference summaries, confirmed the effectiveness of our summarization technique. This head-to-head comparison against a well-established model serves as a strong validation of our research's value and innovation in the field.

Furthermore, our research provided compelling evidence of the limitations of extractive summarization techniques when compared to our abstractive method. In particular, we surpassed the extractive approach outlined in "A Compositional Context Sensitive Multi-Document Summarizer [12]" across a spectrum of ROUGE metrics. This comparison highlights the practical and quantitative advantages of our abstractive approach, which excels in generating coherent and contextually rich summaries. Overall, these results affirm the potency and adaptability of our methodology, positioning it as a promising and substantial advancement in the realm of high-quality abstractive summarization. The implications of our research extend beyond the specific models tested, as our approach provides a template for improved text summarization techniques and opens the door to more effective and contextually relevant summarization in various applications, from news aggregation to information retrieval and natural language understanding.

## References

- [1] Gao, Y., Zhao, W., and Eger, S. SUPERT: Towards New Frontiers in Unsupervised Evaluation Metrics for Multi-Document Summarization. *arXiv preprint arXiv:2005.03724*, 2020.
- [2] Jameel, M.S., Anubhav, Singh, N., Singh, N.K., Singh, C.T., and Ghose, M.K. An Intelligent Automatic Text Summarizer. In *Proceedings of the First International Conference on Intelligent Human Computer Interaction: (IHCI 2009) January 20–23, 2009 Organized by the Indian Institute of Information Technology, Allahabad, India*, Springer India, pp. 223-230, 2009.
- [3] Radev, D.R., Jing, H., Stys, M., and Tam, D. Centroid-Based Summarization of Multiple Documents. *Information Processing & Management*, vol. 40, no. 6, pp. 919-938, 2004.
- [4] Tahseen, R., Omer, U., Farooq, M.S., and Adnan, F. Text Summarization Techniques using Natural Language Processing: A Systematic Literature Review. *VFAST Transactions on Software Engineering*, vol. 9, no. 4, pp. 102-108, 2021.
- [5] Thakur, A., Satish, M. Text Summarizer Using Julia. *International Journal for Research in Applied Science & Engineering Technology*, vol. 10, no. 1, pp. 1371-1375, 2022.
- [6] Widyassari, A.P., Rustad, S., Shidik, G.F., Noersasongko, E., Syukur, A., and Affandy, A. Review of Automatic Text Summarization Techniques & Methods. *Journal of King Saud University-Computer and Information Sciences*, vol. 34, no. 4, pp. 1029-1046, 2022.
- [7] Jalil, Z., Nasir, J.A., and Nasir, M. Extractive Multi-Document Summarization: A Review of Progress in the Last Decade. *IEEE Access*, vol. 9, pp. 130928-130946, 2021.
- [8] Azmi, A.M. and Al-Thanyyan, S. A Text Summarizer for Arabic. *Computer Speech & Language*, vol. 26, no. 4, pp. 260-273, 2012.
- [9] Delort, J.Y., Bouchon-Meunier, B., and Rifqi, M. Enhanced Web Document Summarization using Hyperlinks. In *Proceedings of the fourteenth ACM conference on Hypertext and hypermedia*, pp. 208-215, 2003.
- [10] Hassel, M. and Mazdak, N. FarsiSum-A Persian Text Summarizer. In *Proceedings of the workshop on computational approaches to Arabic script-based languages*, pp. 82-84, 2004.
- [11] Kupiec, J., Pedersen, J., and Chen, F. A Trainable Document Summarizer. In *Proceedings of the 18th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 68-73, 1995.
- [12] Nenkova, A., Vanderwende, L., and McKeown, K. A Compositional Context Sensitive Multi-Document Summarizer: Exploring the Factors That Influence Summarization. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 573-580, 2006.
- [13] Gani, S., Uddin, J., and Mobin, M.I. Extractive Text Summarization Technique using Fuzzy C-Means Clustering Algorithm. In *2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2)*, IEEE, pp. 1-6, 2019.
- [14] Shi, T., Keneshloo, Y., Ramakrishnan, N., and Reddy, C.K. Neural Abstractive Text Summarization with Sequence-to-Sequence Models. *ACM Transactions on Data Science*, vol. 2, no. 1, pp. 1-37, 2021.
- [15] Yeh, J.Y., Ke, H.R., and Yang, W.P. Chinese Text Summarization using a Trainable Summarizer and Latent Semantic Analysis. In *Digital Libraries: People, Knowledge, and Technology: 5th International Conference on Asian Digital Libraries, ICADL 2002 Singapore, December 11–14, 2002 Proceedings 5*, Springer Berlin Heidelberg, pp. 76-87, 2002.
- [16] Ma, C., Zhang, W.E., Guo, M., Wang, H., and Sheng, Q.Z. Multi-Document Summarization via Deep Learning Techniques: A Survey. *ACM Computing Surveys*, vol. 55, no. 5, pp. 1-37, 2022.
- [17] Christian, H., Agus, M.P., and Suhartono, D. Single Document Automatic Text Summarization using Term Frequency-Inverse Document Frequency (TF-IDF). *ComTech: Computer, Mathematics and Engineering Applications*, vol. 7, no. 4, pp. 285-294, 2016.
- [18] Yeh, J.Y., Ke, H.R., Yang, W.P., and Meng, I.H. Text Summarization using a Trainable Summarizer and Latent Semantic Analysis. *Information processing & management*, vol. 41, no. 1, pp. 75-95, 2005.