Tech Science Press

Educational Videos Subtitles' Summarization Using Latent Dirichlet Allocation and Length Enhancement

Sarah S. Alrumiah[*] and Amal A. Al-Shargabi

Department of Information Technology, College of Computer, Qassim University, Buraydah, 51452, Saudi Arabia
[*]Corresponding Author: Sarah S. Alrumiah. Email: saraalrumih@gmail.com

Abstract: Nowadays**,** people use online resources such as educational videos and courses. However, such videos and courses are mostly long and thus, summarizing them will be valuable. The video contents (visual, audio, and subtitles) could be analyzed to generate textual summaries, i.e., notes. Videos' subtitles contain significant information. Therefore, summarizing subtitles is effective to concentrate on the necessary details. Most of the existing studies used Term Frequency–Inverse Document Frequency (TF-IDF) and Latent Semantic Analysis (LSA) models to create lectures' summaries. This study takes another approach and applies Latent Dirichlet Allocation (LDA), which proved its effectiveness in document summarization. Specifically, the proposed LDA summarization model follows three phases. The first phase aims to prepare the subtitle file for modelling by performing some preprocessing steps, such as removing stop words. In the second phase, the LDA model is trained on subtitles to generate the keywords list used to extract important sentences. Whereas in the third phase, a summary is generated based on the keywords list. The generated summaries by LDA were lengthy; thus, a length enhancement method has been proposed. For the evaluation, the authors developed manual summaries of the existing "EDUVSUM" educational videos dataset. The authors compared the generated summaries with the manual-generated outlines using two methods, (i) Recall-Oriented Understudy for Gisting Evaluation (ROUGE) and (ii) human evaluation. The performance of LDA-based generated summaries outperforms the summaries generated by TF-IDF and LSA. Besides reducing the summaries' length, the proposed length enhancement method did improve the summaries' precision rates. Other domains, such as news videos, can apply the proposed method for video summarization.

Keywords: Subtitle summarization; educational videos; topic modelling; LDA; extractive summarization

Abbreviations:

AI	Artificial Intelligence
ASR	Automatic Speech Recognition

BERT	Bidrirectional Encoder Representations from Transformers
CV	Computer Vision
EDUVSUM	The name of an Educational Videos Summaries dataset
IDF	Inverse Document Frequency
LDA	Latent Dirichlet Allocation
LSA	Latent Semantic Analysis
ML	Machine Learning
MMLDA	MultiModal LDA
NLP	Natural Language Processing
NTM	Neural Topic Model
ROUGE	Recall-Oriented Understudy for Gisting Evaluation
ROUGE-1	ROUGE measurement that measures the overlap of unigrams
ROUGE-2	ROUGE measurement that measures the overlap of bigrams
ROUGE-L	ROUGE measurement that measures the longest matching sentences
TF-IDF	Term Frequency–Inverse Document Frequency

1 Introduction

With the availability of online learning, i.e., learning through the internet [1], educational videos' generation increased. Educational videos vary in duration, content, and presentation styles. For instance, lectures' videos usually present the subject's PowerPoint slides and have long durations (more than 30 mins). In contrast, kids' educational videos are often short (1 to 5 mins) with attractive animations. Even though long educational videos contain great information, people usually avoid watching them [2]. Searching for a piece of knowledge in a long video is challenging, as it requires lots of time. Thus, there is a need for techniques to summarize such videos.

Summarizing educational videos benefits learners by saving their time and storage space and making the searching and indexing process quick and easy [3–5]. A video can be summarized based on its audio [3,5–7], visual [3,5,7,8], and textual [3,5–14] data, i.e., subtitles. The video's audio is summarized by converting the spoken words to text using speech-to-text techniques and applying text-based summarization methods [6]. Besides, summarizing a video based on its visual content to a textual form using different techniques, such as video and image tagging and captioning [15,16]. Whereas in subtitles summarization, the textual data is summarized using text-based summarization algorithms [9]. On the other hand, the generated summaries could be either a short video [17] or a text [3–14,16,18].

In textual summarization, the input (need to be summarized) and output (generated summary) are in textual form. The textual outlines can be classified based on the number of input documents to a single document and multi-documents [19]. Moreover, textual summarization is classified based on the used algorithm for abstractive and extractive summarization [9]. The abstractive summarization summarizes a document like human summarization using external vocabulary and paraphrasing [4]. In contrast, extractive summarization summarizes a document by extracting the exact sentences that are considered significant by calculating each sentence's frequency and importance score [3,5–14]. Tab. 1 compares the two summarization techniques.

This paper aims to summarize educational videos to save learners' time and resources and provide quick and straightforward searching and indexing processes. The main focus is on summarizing subtitles because in most of the lectures, the visual content, e.g., slides, are pre-given to students and most educational websites provide videos' transcripts. Therefore, in lectures' videos, the significant concentration is on the spoken sentences present in the video's subtitles and transcripts. This study uses subtitle files available in the "EDUVSUM" educational videos'

dataset [17]. Extractive text summarization helps extract valuable sentences. Therefore, single document extractive summarizations are applied to videos' subtitles. Regarding the scientific content of courses and lectures, people need exact sentences without paraphrasing. Additionally, the current work focuses on generating text summaries as most students prefer referring to the lecture's textual notes [20]. Fig. 1 illustrates the scope of this work highlighted in blue.

Table 1: A comparison between abstractive and extractive summarization

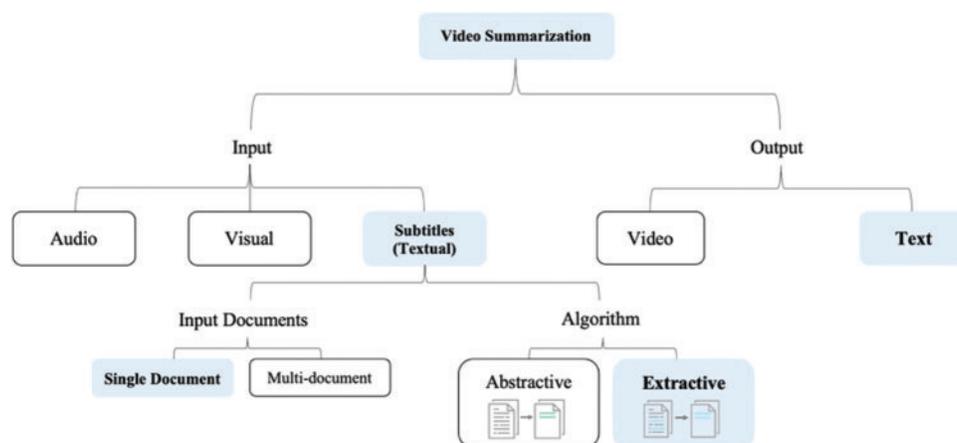| Abstractive summarization	Extractive summarization
Uses external vocabulary	Uses in-text vocabulary only
Word-level attention	Sentence-level attention
Requires complex algorithms and rules	Requires ranking algorithms
Language-dependent summaries	Language-independent summaries

Figure 1: The scope of the study

Based on the literature, we can derive the following observations:

(1) Educational videos' datasets are limited and not sufficient for summarization purposes. The available datasets mostly include short clips and miss subtitles or transcripts and require major manual preprocessing. Moreover, to the best of our knowledge, educational videos' datasets lack the human-generated summaries necessary for the model evaluation process.

(2) Although Latent Dirichlet Allocation (LDA) recorded the highest performance in summarizing documents, it was not applied in videos' subtitles summarization. However, different algorithms were applied. These include Latent Semantic Analysis (LSA) [3,7], Term Frequency–Inverse Document Frequency (TF-IDF) [9,10], and Bidirectional Encoder Representations from Transformers (BERT) [11,13,17].

(3) LDA-based document summaries always tend to be lengthy [5].

Thus, this work proposes the use of LDA to summarize educational videos based on their subtitles. The main contributions of this study are:

(1) Extending the "EDUVSUM" educational videos' dataset [17] by generating human summaries from subtitle files, as the original dataset only includes videos and subtitles; without human summaries.
(2) A proposed summarization method based on LDA.
(3) A proposed method for enhancing summaries in terms of length and quality.

The two proposed methods mentioned above were validated using experiments.

This paper is structured as follows: Section 2 illustrates some of the related works. Section 3 presents the materials and methodology. Section 4 shows the results, Section 5 discusses the study's outcomes and Section 6 concludes the study.

2 Related Work

Many researchers have applied data mining techniques to enhance the education field, such as analyzing students' performance [21] and studying the discoursal identity in postgraduates' academic writings [22]. On the other hand, the scope of this study focuses on the extractive summarization of educational videos. Extractive summarization extracts valuable information without modification [23]. For instance, extractive text summarization summarizes a document by selecting the important sentences. Extractive text summarization starts with accepting an input document and preprocess the document, such as removing punctuations and stop-words [19]. Then feature extraction, e.g., a bag of words, is applied to the preprocessed text. After that, sentences are represented using vector representation as an example. Finally, based on some algorithms and criteria, sentence selection is applied to generate the summary. Fig. 2 summarizes the extractive text summarization process.

Figure 2: Extractive text summarization process

This section discusses the previous work done to summarize educational videos, which can be divided into three categories (i) video summarization based on audio, visual scenes, and subtitles [3,5,7,8,16–18], (ii) audio-only outlines [6], and (iii) subtitles-only summarization [3,5–14]. Speech recognition faces some challenges when generating text summaries [6], e.g., lack of punctuations [9]. This study focuses only on the videos' and subtitles' summarization due to audio-related issues and the availability of online subtitles' generation tools. Moreover, subtitles are considered as documents. Therefore, this section discusses some of the document extractive summarization efforts. The works discussed here were selected based on their relevancy to the study's scope.

2.1 Summarization Based on Audio, Visual, and Subtitles

Aggregating the Natural Language Processing (NLP), Automatic Speech Recognition (ASR), and Computer Vision (CV) techniques assisted researchers in generating video summaries [5]. This section discusses some of the applied video summarization methods.

In [17], the authors analyzed audio, visual, and textual contents of 98 educational videos to assign scores to important video segments. The authors used a python-based method to extract audio features. Moreover, researchers used Xception, ResNet-50, VGG-16 and Inception-v3 to

extract the visual features and BERT for textual features. An annotation tool and EDUVSUM dataset, i.e., a dataset of annotated educational videos, were generated. Researchers concluded that visual features are not well suited for the academic domain.

Additionally, multimedia news was summarized using Multimodal LDA (MMLDA) [5]. Furthermore, video summarization has been applied to movies [8]. Researchers in [8] summarized 156 scenes of the Forrest Gump movie based on scenes' description and subtitles analysis. The authors found that subtitles have a positive effect on generating summaries.

Other studies focused on summarizing long videos to short visual scenes and textual summaries [7]. Researchers applied Luhn, LSA, LexRank, and TextRank algorithms to evaluate the best algorithm by summarizing one-hour long videos. After assessing the generated summaries with Recall-Oriented Understudy for Gisting Evaluation (ROUGE) and human comparison, the authors recorded that LSA and LexRank provided the best results compared with other methods.

Authors in [3] created textual video summaries by first generating subtitles using speech recognition techniques, then applying text summarization algorithms based on NLP. Like [7], researchers in [3] used Luhn, LSA, LexRank, and TextRank to generate textual summaries. Their results agreed with [7] that LSA performed well and had the best contribution.

Learning how to directly map a sequence of video frames, i.e., scenes, to a series of words to generate video captions was studied in [16]. As a result, developing video descriptions is challenging as it is difficult to determine the salient content and describe it appropriately.

On the other hand, authors in [18] generated fixed-length textual summaries from asynchronous videos, i.e., videos without descriptions and subtitles. Researchers used LexRank and Neural Topic Model (NTM) to produce textual summaries of five to ten minutes' news videos. Researchers reported that videos transcripts are essential to generate summaries, whereas audio and visual contents have a limited effect on summarization performance.

2.2 Summarization Based on Subtitles Only

Furthermore, some studies focused on subtitles summarization to generate textual summaries of educational videos. This section presents some of the applied subtitles summarization methods.

Researchers in [9] applied lectures' subtitles summarization of fixed-length sentences to avoid the miss-identified punctuation marks in the speech-to-text summaries. TF-IDF was used to find meaningful sentences. The authors concluded that outline is effective when punctuations are part of subtitles.

Another research that produced lectures videos summarization by analyzing subtitles using TF-IDF [10]. Authors in [10] treated each sentence as a document and generated a summary of sentences with a threshold equal to the average TF-IDF value of all sentences. Thus, based on the IDF relevancy term, the lesser the term's occurrence, the higher its importance. In conclusion, researchers found that extractive text summarization reduced the original content by 60%. Further, removing stop-words does not affect the produced summary.

Moreover, a cloud-based lecture summarization service was generated [11]. Researchers used BERT to generate lectures subtitles summaries. Besides the summarization service, the system also provided storage and management services. K-mean cluster was used to assist in the summary selection by identifying the closest sentence to the centroid. An extended work of [11] is presented in [13], where authors added a dynamic method to select the appropriate number of clusters besides using BERT to produce summaries. Depending on the size of the cluster, the generated

summary overcame the drawback of generating short summaries. However, most of the human-generated outlines contained three to four sentences which cannot be taken as a standard for dynamically predicting the number of sentences in summary.

As subtitles consider as documents when analyzing them, the following paragraphs illustrate the extractive document summarization works. Document summarization is classified based on the number of input documents to a single document and multi-documents.

Authors in [12] generated the MUSEEC tool to summarize documents. MUSEEC is an extractive text summarization tool that consists of MUSE, POLY, and WECOM methods. MUSE provided supervised learning extractive summarization, while POLY produced unsupervised learning summaries. WECOM was used to shorten sentences. Furthermore, MUSEEC is a language-independent summarization tool. Thus, MUSEEC has been used and improved in other studies, such as movie summarization [8].

Additionally, others combined the power of topic modelling with the simplicity of extractive summarization to produce document summaries [14]. LDA proved its effectiveness in generating summaries as it improved the TF-IDF results. Nevertheless, using topic modelling induce the need for pre-determined topic specifications. Furthermore, authors in [24] combined topic modelling using LDA with classification techniques to generate documents summaries. Topic modelling-based document summarization deals with some challenges, such as out of control output and the possibility of missing some expected topics [5]. However, summarizing documents with topic modelling algorithms generated good summaries.

From the literature, we can see that LSA recorded high results when applied to videos summarization. In comparison, studies that focused on subtitles-based outlines used TF-IDF and BERT only. Moreover, although LDA proved its effectiveness in summarizing documents, it has not been applied to subtitles summarization. Regardless of structural differences, both a document file and a subtitle file contain a collection of sentences. Therefore, this work aims to apply topic modelling using LDA on educational videos subtitles to generate lectures' summaries.

3 Materials and Methods

This section discusses the materials and methods used to obtain the study's results. The proposed LDA-based subtitles summarization model and dataset details are presented, as well as the proposed summaries' length enhancement method and the evaluation methods. Fig. 3 illustrates the followed methodology phases.
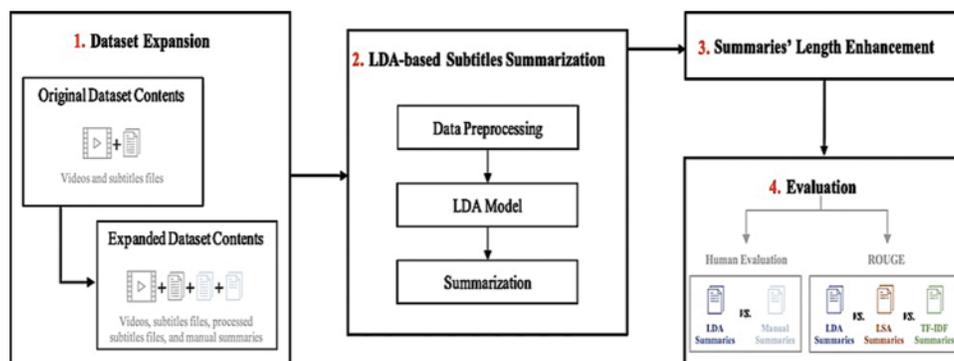

Figure 3: The study's methodology phases

3.1 Dataset Expansion

This work used the educational videos' subtitle files from EDUVSUM [17] dataset. EDU-VSUM is an English educational videos dataset containing 98 MP4 videos with 98 subtitles files of each video [17]. All videos presented in the dataset are in English. The videos describe topics in various fields, such as computer science, chemistry, and biology. Additionally, all videos have a duration of fewer than 16 min; see Tab. 2 for more information about videos durations.

Table 2: EDUVSUM datasets videos count along with their duration

Duration (minutes)	≥ 14	13	12	11	10	9	8	7	6	≤ 5
Count of videos (videos)	2	2	10	5	7	7	9	6	3	47

Considering ten minutes length videos as short educational videos, authors excluded less than ten minutes long videos. Moreover, the authors developed manual summaries for the selected 26 videos, as the used dataset, i.e., EDUVSUM, did not provide subtitles' summaries. Authors created summaries' files by selecting the valuable sentences in the original subtitles' files. The manual-generated outlines were about 50-55% of the original subtitles' files, i.e., the original subtitles' contents were reduced to half in the manual summaries. On the other hand, the dataset's subtitle files' structure needed some manual preprocessing and modification due to using an online tool in generating the existing subtitles in [17]. Therefore, the authors removed time ranges and merged the sentences manually. During the experiments, the authors used a sample of five videos out of the selected 26 videos.

3.2 LDA-Based Subtitles Summarization Model

Summarizing subtitles is similar to documents summarization, where the focus is on valuable information, i.e., sentence, to add it in the summary. In this paper, the use of subtitles and documents is interchangeable. An input document (d) is a set of (n) number of sentences (s), $d = \{s_1, s_2, \ldots, s_n\}$. Moreover, a sentence is a collection of words (w), $s = \{w_1, w_2, \ldots, w_n\}$. To summarize a subtitle file, the authors extracted the informative sentences from the original document using a topic modelling technique. Topic modelling is an unsupervised learning approach that aims to organize and classify textual data by discovering words' patterns [25].

Additionally, identifying keywords, i.e., topics, help in determining valuable sentences. For instance, in an Artificial Intelligence (AI) lecture, the essential information could contain keywords, such as AI, Machine Learning (ML), supervised, unsupervised, modelling, etc. To extract the crucial words, the authors used LDA. LDA is a topic modelling method that uses a statistical approach to discover important topics by analyzing the document's words [26].

LDA is a Bayesian-based model that decomposes a document into a set of topics [26]. LDA uses a vector of random multinomial parameters (θ) in (n) documents. The distribution of θ is affected by hyperparameters, α and B matrix. B matrix illustrates the relationship between the document's discrete topic variables z_{ij} and words w_{ij}. Eq. (1) shows the LDA probability of a list of words (W) [26].

$$P(W|\alpha, B) = \prod_{i=1}^{n} \int^P (\theta_i|\alpha) \left[\prod_{j=1}^{m_n} P(w_{ij}|\theta_i, B) \right] d\theta_i \tag{1}$$

The probability of a set of recognized words (W) depends on α and B matrix values. Moreover, d is an index of a document with (m) words. Fig. 4 presents a graphical LDA model [26].
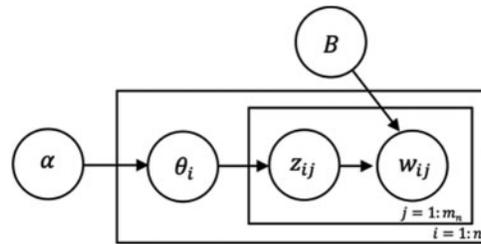

Figure 4: Graphical representation of the proposed LDA model

Fig. 5 presents the proposed LDA summarization model framework. The subtitles' summarization process starts with acquiring a single subtitle file. Then preprocess the subtitle file, i.e., removing punctuations, converting letters to lowercase, splitting the document into sentences, removing stop words, and creating an id2word dictionary and corpus. The corpus, id2word dictionary, and the number of topics (in our case, one topic) are passed to the LDA model to generate the topic's keywords list. The keywords list contains ten words considered the most important words when extracting and selecting the summary's sentences. Any sentence that includes one of the ten words is added to the output summary. Then, the generated summary is compared with the human outlines using some evaluation methods.
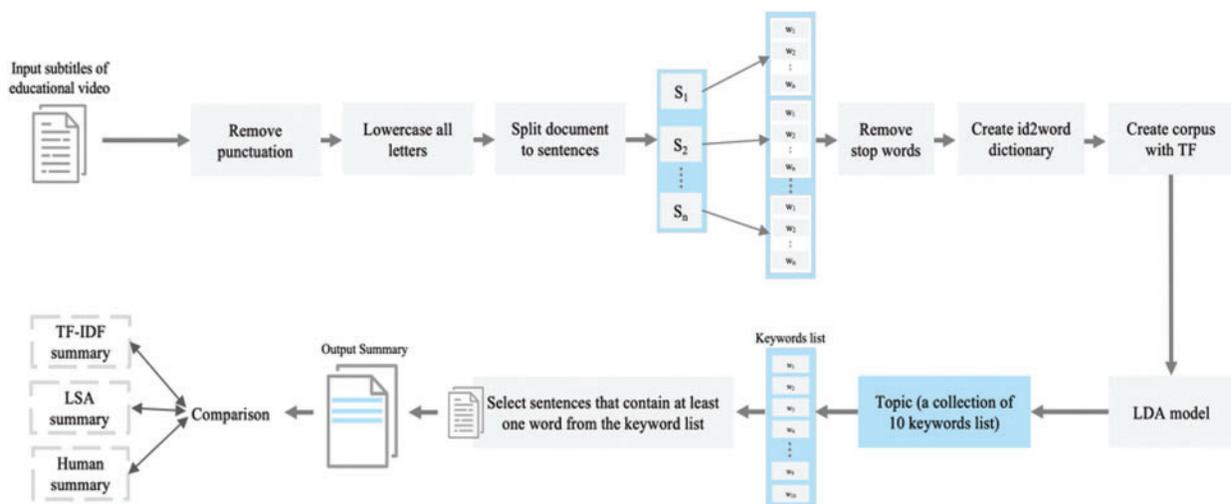
Figure 5: LDA-based subtitle summarization model framework

This work developed a single document summarization model. Each subtitle file is treated as a single document. Additionally, to identify the best number of topics, the authors used Grid Search, an optimization algorithm besides the Scikit-learn LDA model [27]. The grid search results with the best estimator of one topic, which is reasonable as a lecture usually describe a specific

topic. Therefore, the summarization model accepts a single subtitle file and generates a single topic that describes the lecture's contents.

The Scikit-learn's based LDA model does not link words with the generated output; only topic statistics [27]. Thus, the authors implemented a Gensim-based LDA model. Gensim is a topic modelling Python library that is human-readable as it outputs the generated topics [28]. One topic in Gensim contains ten words. Those words are used as keywords to extract sentences to appear in the generated summary.

3.3 Summaries' Length Enhancement

As LDA dynamically generates topics that differ from one run to another, the output summary's length varies and cannot be controlled. Moreover, LDA tends to generate lengthy summaries. Thus, the authors implemented a method to reduce the number of selected sentences by reducing the keywords list. The keywords list contains the ten words that the Gensim LDA model generated. To minimize the keywords list, non-noun words, e.g., verbs, number words, etc., were removed.

3.4 Model Evaluation

The LDA-based generated summaries are compared with manual-generated outlines. Moreover, TF-IDF and LSA models were implemented to compare their performance with the proposed LDA model on the same subtitles' files. TF-IDF determines the relative words in a document by calculating the term frequency and inverse document frequency score for each word [29]. The sentence importance score is calculated based on the terms' scores of that sentence. Contrarily, LSA is a method that extracts the meanings and semantics of words in a document [30]. A Sumy-based LSA model was used. Sumy is a Python package for extractive text summarization with a flexible feature of specifying the number of generated sentences [31].

ROUGE and human evaluation were used to evaluate the LDA-based generated summaries. ROUGE assesses the generated summaries by comparing them with the human-generated summaries. ROUGE contains some measurements, such as ROUGE-1, ROUGE-2, and ROUGE-L. For instance, ROUGE-1 measures the overlap of unigrams, while ROUGE-2 measures bigrams overlap. Additionally, ROUGE-L measures the longest matching sentences. However, ROUGE mainly focuses on comparing word sequences; therefore, human assessment is needed to evaluate sentences that existed in both the generated summary and the human summary. Human evaluation is done by calculating the number of sentences in summaries that matched the human outline, dividing it by the total number of sentences in the generated summary.

4 Results

This section presents the experimental results of the subtitles summarization. The summarization process consisted of three phases, (i) dataset expansion phase, (ii) LDA-based subtitles summarization phase, and (iii) summaries' length enhancement phase. The following sections illustrate the results regarding each phase.

4.1 Dataset Expansion Phase

As mentioned in Section 3.1, the subtitles files needed manual preprocessing before being summarized. Therefore, the subtitles files were manually preprocessed by removing time ranges and merging sentences, as shown in Fig. 6. On the other hand, the authors generated manual summaries for evaluation purposes. As a result, the current expanded version of the EDUVSUM

dataset includes 26 processed subtitle files and 26 manual summaries along with the original videos and subtitle files.
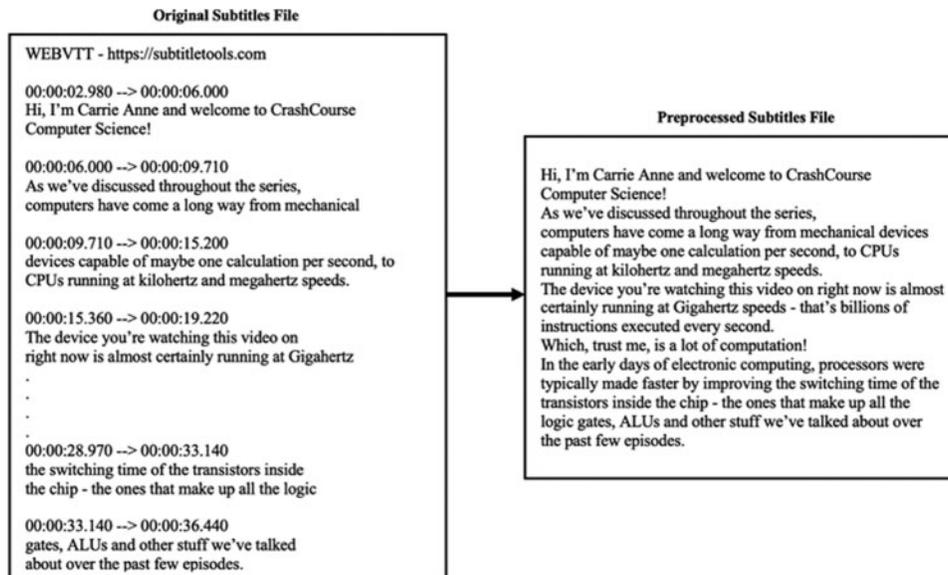

Figure 6: Sample of a subtitle file portion in the EDUVSUM dataset and the preprocessed version

4.2 LDA-Based Subtitles Summarization Phase

After obtaining processed subtitle files in the dataset expansion phase, the subtitle files are ready to be summarized. To summarize subtitles using the proposed LDA approach, the authors implemented a Python-based script. The summarization phase started with obtaining a subtitle file as an input. Then preprocess the subtitle file for summarization by lowercase all letters and removing punctuations. Additionally, to prepare the data for LDA, authors split the subtitle file into sentences and each sentence is divided into its word creating a list of sentences' words. After that, an id2word dictionary is created that contained identification numbers for each word. Then, a corpus of sentences is generated. The corpus represented the term frequency for each word in a sentence using the word's id. The Gensim LDA model is then generated by passing the corpus, id2word dictionary, and the number of topics, i.e., based on Section 3.2, the num_topics = 1. The Gensim-based LDA represent the topic by outputting a list of 10 keywords, see Fig. 7 (note that word cloud was used for visualization purposes only). The authors extracted the ten keywords and used them to select the sentences included in the summary, as shown in Fig. 8.

After developing subtitles' summaries with the LDA model and comparing LDA performance with TF-IDF and LSA (see Fig. 9), Tab. 3 describes the ROUGE and manual evaluation results. Tab. 3 shows that LDA-based summaries recorded the highest ROUGE scores and human evaluation as well. Moreover, the number of sentences in the generated summaries is critical in the evaluating process. Therefore, Tab. 4 shows the average number of sentences in the TF-IDF, LSA, and LDA summaries compared with human-generated summaries and the mean sentences in the original subtitles' files. LSA-based summaries produced the exact number of sentences in the human-generated summaries due to the LSA model's flexibility in controlling the number

of sentences in the output summaries. In comparison, TF-IDF developed the shortest outlines. However, LDA-based summaries included the highest number of sentences.
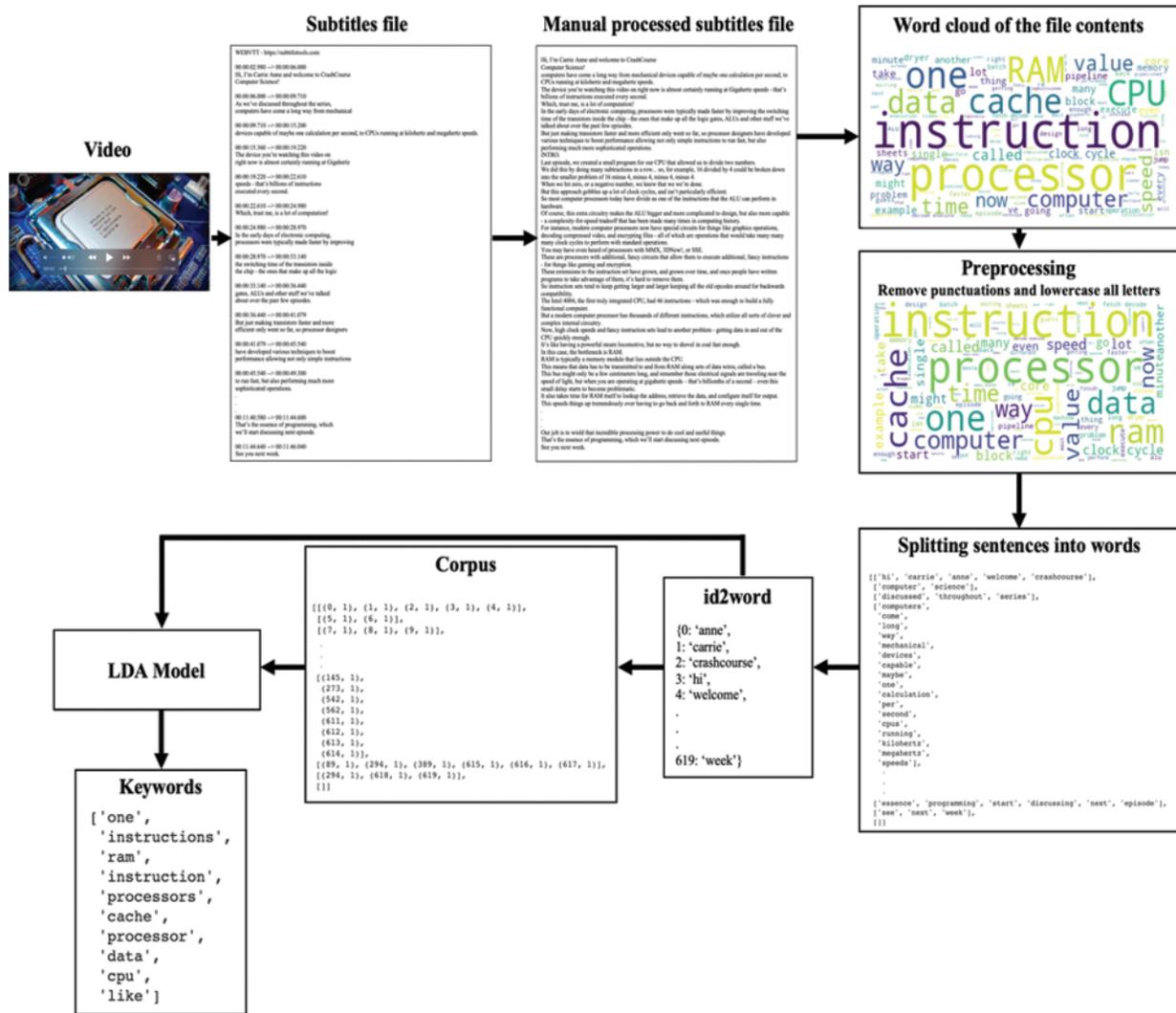
Figure 7: Sample of the LDA-based subtitles summarization steps

4.3 Summaries' Length Enhancement Phase

To enhance LDA-based lengthy summaries, the authors excluded non-noun words from the keywords list, as explained in Section 3.3 and Figs. 10 and 11. Tab. 5 compares the ROUGE scores and human evaluation of LDA summaries before and after applying the length enhancement method. The precision rate increased in the enhanced LDA summaries as the sentences' length decreased, as shown in Tab. 6.

Figure 8: Sample of the LDA-based subtitles summarization output

5 Discussion

Based on our results, the performance of LDA-based subtitles summarization surpassed the existing TF-IDF and LSA models. As LDA generated a keywords list of the lecture's topic, the LDA-based summarization model focused on sentences that contain the lecture's important words. However, LDA summaries were the longest in terms of sentences. A large number of sentences in the generated summaries could affect ROUGE scores. Thus, the authors tried to enhance summaries' length by eliminating non-noun words from the LD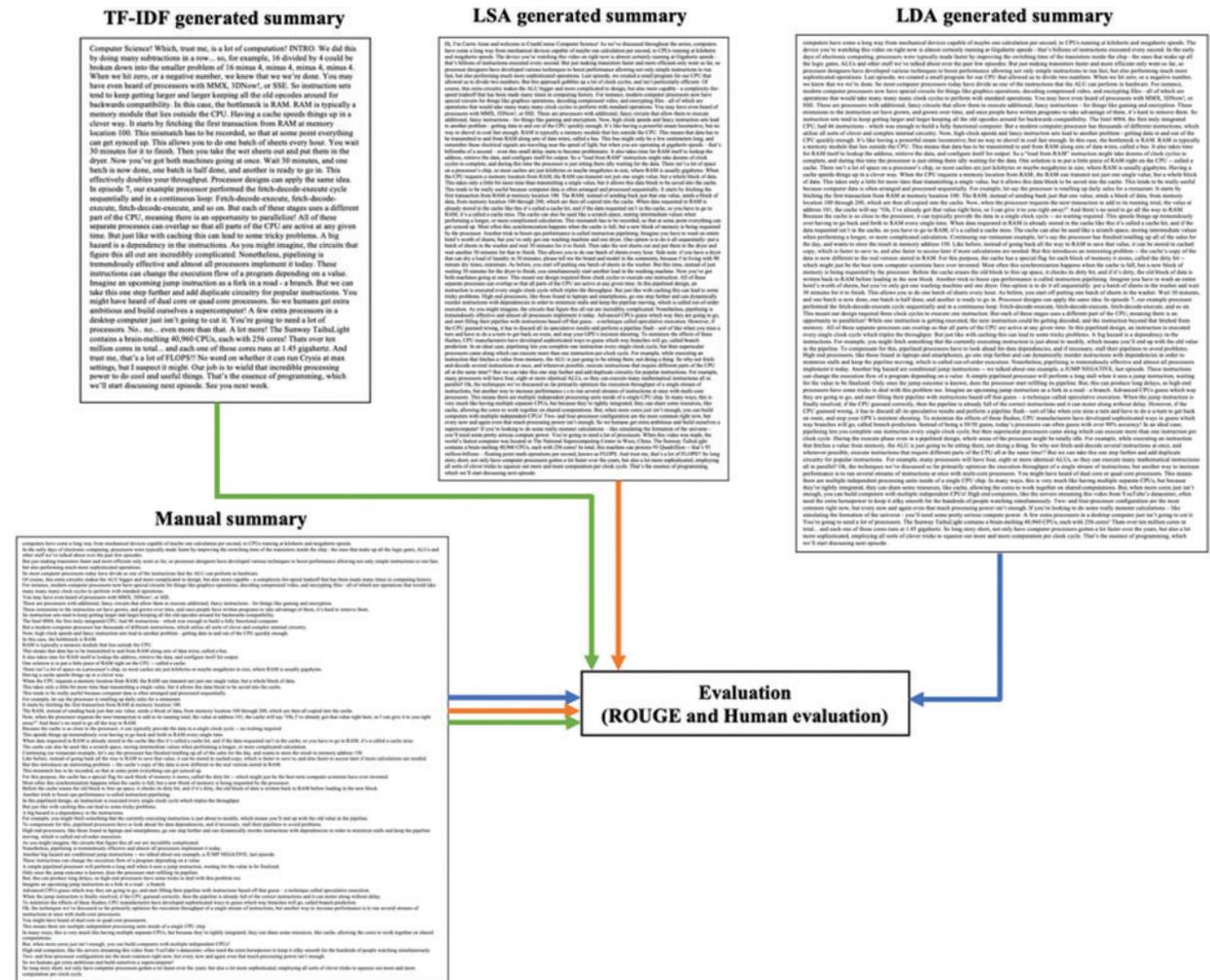A generated keywords list. The length enhancement method improved the precision performance; as the number of non-relevant sentences reduces, the precision rate increases. However, the length enhancement approach based on nouns may not be suitable for topics that majorly consider numbers, Booleans, and verbs.

On the other hand, controlling the length of generated summaries in TF-IDF and LDA is challenging. In contrast, the flexibility of the LSA model resulted in summaries with a number of sentences equal to the number of sentences in the human-generated summaries. Moreover, using TF-IDF to summarize educational content is insufficient because TF-IDF focuses on the less appearing words and sentences. Where in lectures, the most repeated words are considered necessary. Furthermore, the enhanced LDA summaries overpass LSA-based summaries in terms of ROUGE scores and human evaluation. To sum up, LDA proved its effectiveness in summarizing educational subtitles' files.

Figure 9: Sample of the subtitles summaries evaluation

Table 3: ROUGE scores and human evaluation of subtitles' summarization

Model	ROUGE-1			ROUGE-2			ROUGE-L			Human Evaluation
	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	
TF-IDF	0.3688	0.6966	0.2524	0.2194	0.4088	0.1510	0.3806	0.5838	0.2846	0.3669± 0.0967
LSA	0.7156	**0.6980**	0.7380	0.5590	0.5456	0.5760	0.6796	0.6442	0.7262	0.5610± 0.0619
LDA	**0.7818**	0.6964	**0.8992**	**0.7132**	**0.6336**	**0.8228**	**0.7814**	**0.7166**	**0.8666**	**0.6216± 0.0361**

Table 4: The average number of sentences in the generated summaries

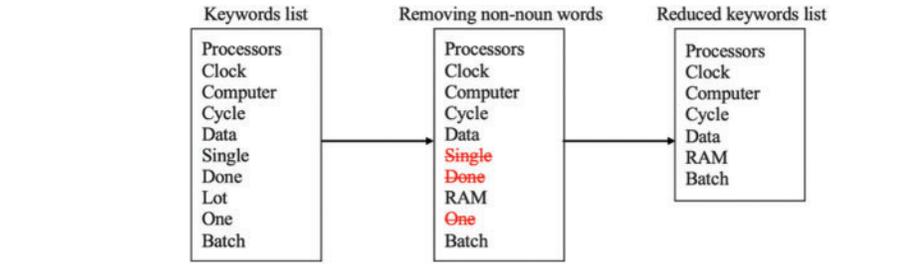| Average number of sentences				
Generated summaries			Human-generated summaries	Original subtitles'
TF-IDF	LSA	LDA		
41.6	**62.6**	81.4	**62.6**	120.6

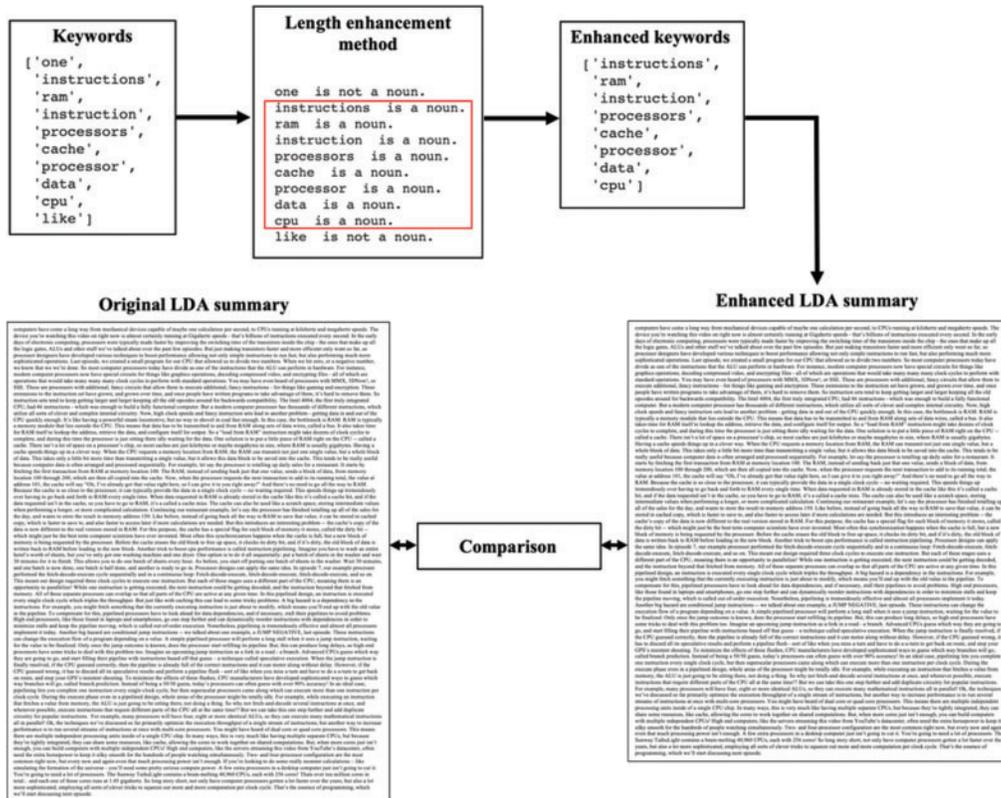**Figure 10:** Sample of LDA keywords list reduction

Figure 11: Sample of the length enhancement process

Table 5: LDA subtitles' summarization results with length enhancement

Summaries	ROUGE-1			ROUGE-2			ROUGE-L			Human evaluation
	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	
Original LDA	0.7818	0.6964	0.8992	0.7132	0.6336	0.8228	0.7814	0.7166	0.8666	0.6216± 0.0361
Length's enhancement LDA (only nouns keywords)	0.7650	**0.7544**	0.7912	0.6672	**0.6532**	0.6938	0.7324	**0.7306**	0.7450	**0.6364± 0.0723**

Table 6: The average number of sentences in LDA-based summaries with length enhancement

Average number of sentences	
Original LDA	Enhanced LDA
81.4	**65.6**

6 Concluding Remarks

Learners spent a lot of their time watching long educational and lecture videos. Summarizing long videos in textual form can be effective. Thus, to increase learning effectiveness and reduce the learning time, the authors implemented an LDA-based subtitles summarization model. The subtitles of educational videos were summarized using an LDA-generated keywords list. Regarding the results, LDA recorded the highest performance compared with LSA and TF-IDF summarization models. Furthermore, reducing LDA summaries' length by extracting non-noun words from the keywords list did improve the LDA precision rate and human evaluation.

Students in any field can use the proposed work to generate lectures' summaries. Moreover, the authors encourage interested researchers to consider applying document-based analysis in videos' textual contents. The proposed model could be applied in other domains, such as news videos. Further, in the educational context, the type and contents of a lecture could affect the generated summaries, e.g., a quiz or test-related details are essential even though they are out of the lecture's topic scope. On the other side, based on personal perspectives, the human-generated summaries in this work could differ from one to another. Therefore, the model-based summarization performance could be affected.

Nevertheless, controlling the length of LDA-based generated summaries could be considered in the future. Moreover, working on punctuations and nouns' properties, e.g., singular and plural, could be a future improvement aspect. Additionally, in the educational domain, repeated words could be important.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] R. D. Johnson and K. G. Brown, "E-Learning," In *The Wiley Blackwell Handbook of the Psychology of the Internet at Work*, 1st ed., Hoboken, NJ, USA: John Wiley & Sons, vol. 17, pp. 369–400, 2017.

[2] M. Carmichael, A. K. Reid and J. D. Karpicke, "Assessing the Impact of Educational Video on Student Engagement, Critical Thinking and Learning: *The Current State of Play*, London, England, UK: SAGE, pp. 1–22, 2018. [Online]. Available: https://us.sagepub.com/sites/default/files/hevideolearning.pdf.

[3] V. B. Aswin, M. Javed, P. Parihar, K. Aswanth, C. R. Druval *et al.,* "NLP-driven ensemble-based automatic subtitle generation and semantic video summarization technique," in *Advances in Intelligent Systems & Computing*, vol. 1133, Singapore: Springer, pp. 3–13, 2021.

[4] A. Dilawari and K. M. U. Ghani, "Asovs: Abstractive summarization of video sequences," *IEEE Access*, vol. 7, no. 2019, pp. 29253–29263, 2019.

[5] S. Vazarkar and T. Manjusha, "Video to text summarization system using multimodal LDA," *Journal of Seybold*, vol. 15, no. 9, pp. 3517–3523, 2020.

[6] A. Vinnarasu and V. J. Deepa, "Speech to text conversion and summarization for effective understanding and documentation," *International Journal of Electrical & Computer Engineering (IJECE)*, vol. 9, no. 5, pp. 3642–3648, 2019.

[7] S. Sah, S. Kulhare, A. Gray, S. Venugopalan, E. Prud'hommeaux *et al.,* "Semantic text summarization of long videos," in *IEEE Winter Conference on Applications of Computer Vision, (WACV)*, Santa Rosa, CA, USA, pp. 989–997, 2017.

[8] C. Liu, L. Mark and S. Armin, "Towards automatic textual summarization of movies," in *Recent Developments & the New Direction in Soft-Computing Foundations & Applications*, vol. 393, Baku, Azerbaijan, Cham: Springer, pp. 481–491, 2020.

[9] R. K. Abhilash, A. Choudhary, A. Vaka and D. Uma, "Lecture video summarization using subtitles," in *the 2nd EAI Int. Conf. on Big Data Innovation for Sustainable Cognitive Computing. Proc.: EAI/Springer Innovations in Communication and Computing.*, Coimbatore, India, pp. 83–92, 2020.

[10] S. Garg, "Automatic text summarization of video lectures using subtitles," in *Advances in Intelligent Systems and Computing*, vol. 555, Singapore: Springer, pp. 45–52, 2017.

[11] D. Miller, "Leveraging BERT for extractive text summarization on lectures," ArXiv Preprint, vol. 1, no. 04165, pp. 1–7, 2019.

[12] L. Marina, V. Natalia, L. Mark and C. Elena, "Museec: a multilingual text summarization tool," in *Proc. of ACL-2016 System Demonstrations*, Berlin, Germany, pp. 73–78, 2016.

[13] A. Srikanth, A. S. Umasankar, S. Thanu and S. J. Nirmala, "Extractive text summarization using dynamic clustering and co-reference on BERT," in *5th Int. Conf. on Computing, Communication and Security (ICCCS)*, Patna, India, pp. 1–5, 2020.

[14] N. Gialitsis, N. Pittaras and P. A. Stamatopoulos, "Topic-based sentence representation for extractive text summarization," in *the Recent Advances in Natural Language Processing (RANLP 2019) Conference. Proc. of the Multiling 2019 Workshop*, Varna, Bulgaria, pp. 26–34, 2019.

[15] E. G. Özer, İN Karapınar, S. Başbuğ, S. Turan, A. Utku *et al.,* "Deep learning based, a new model for video captioning," *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 11, no. 3, pp. 514–519, 2020.

[16] S. Venugopalan, M. Rohrbach, J. Donahue, R. Mooney, T. Darrell *et al.,* "Sequence to sequence-video to text," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Santiago, Chile, pp. 4534–4542, 2015.

[17] J. A. Ghauri, S. Hakimov and R. Ewerth, "Classification of important segments in educational videos using multimodal features," in *the 29th ACM International Conference on Information and Knowledge Management (CIKM 2020). Proceedings of the CIKM 2020 Workshops*, Galway, Ireland, pp. 1–8, 2020.

[18] H. Li, J. Zhu, C. Ma, J. Zhang and C. Zong, "Read, watch, listen, and summarize: Multi-modal summarization for asynchronous text, image, audio and video," *IEEE Transactions on Knowledge & Data Engineering*, vol. 31, no. 5, pp. 996–1009, 2019.

[19] P. G. Magdum and S. Rathi, "A survey on deep learning-based automatic text summarization models," in *Advances in Intelligent Systems & Computing*, vol. 1133, Singapore: Springer, pp. 377–392, 2021.

[20] L. Luo, K. A. Kiewra, A. E. Flanigan and M. S. Peteranetz, "Laptop versus longhand note taking: Effects on lecture notes and achievement," *Instructional Science*, vol. 46, no. 2018, pp. 947–971, 2018.

[21] K. Shaukat, I. Nawaz, S. Aslam, S. Zaheer and U. Shaukat, "Student's performance in the context of data mining," in *2016 19th International Multi-Topic Conference (INMIC)*, Islamabad, Pakistan, pp. 1–8, 2016.

[22] N. Kanwal, S. A. Qadir and K. Shaukat, "Writing instructions at a university and identity issues: A systemic functional linguistics perspective," *International Journal of Emerging Technologies in Learning*, vol. 16, no. 6, pp. 275–285, 2021.

[23] N. Moratanch and S. Chitrakala, "A survey on extractive text summarization," in *2017 International Conference on Computer, Communication and Signal Processing (ICCCSP)*, Chennai, India, pp. 1–6, 2017.

[24] R. K. Roul, "Topic modeling combined with classification technique for extractive multi-document text summarization," *Soft Computing*, vol. 25, no. 2021, pp. 1113–1127, 2021.

[25] S. Likhitha, B. S. Harish and H. M. K. Kumar, "A detailed survey on topic modeling for document and short text data," *International Journal of Computer Applications*, vol. 178, no. 39, pp. 1–9, 2019.

[26] I. H. Witten, E. Frank, M. A. Hall and C. J. Pal, "Probabilistic Methods," in *Data Mining Practical Machine Learning Tools and Techniques*, 4th ed., Cambridge, MA, USA: Morgan Kaufmann, vol. 9, pp. 335–416, 2017.

[27] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion *et al.,* "Scikit-learn: Machine learning in python," *Journal of Machine Learning Research*, vol. 12, no. 2011, pp. 2825–2830, 2011.

[28] R. Řehůřek and P. Sojka, "Software framework for topic modelling with large corpora, " in *New Challenges for NLP Frameworks Conference. Proceedings of the LREC 2010 Workshop*, Malta, pp. 45–50, 2010.

[29] J. Bell, "Machine learning with text documents," in *Machine Learning: Hands-on for Developers and Technical Professionals*, 2nd ed., Indianapolis, IN, USA: John Wiley & Sons, vol. 10, pp. 197–221, 2020.

[30] T. K. Landauer, D. S. McNamara, S. Dennis and W. Kintsch, "Introduction to LSA: Theory and Methods," in *Handbook of Latent Semantic Analysis*, 2nd ed., Mahwah, NJ, USA: Psychology Press, vol. 1, pp. 3–71, 2013.

[31] M. Belica, "Sumy: Automatic text summarizer," Accessed 11-july-2021, 2019. [Online]. Available: https://pypi.org/project/sumy/.