ISSN 2007-9737

Edu Vault: An Interactive, Multilingual, and Intelligent Topic-Conscious Video Discovery System for Enhanced Conceptual Learning Using Advanced NLP Techniques

Lopa Mandal[1,*], T. Bhaskara Harsha Vardhan[1], K. Ganesh Narasimha Reddy[1], D. Sai Yeswanth Reddy[1], Sauvik Bal[2,3]

[1] Department of Computer Science and Engineering, Alliance University, Karnataka, India

[2] Department of Computer Science and Engineering, Techno India University, India

[3] Maulana Abul Kalam Azad University of Technology, India

{drmandal.lopa, bhaskaraharshavardhan456, kollaganeshreddy777, yyeswanthreddy526, sauvikbal}@gmail.com

Abstract. The present work developed an intelligent topic-conscious video discovery system to retrieve videos from YouTube to enhance e-learning. Speech recognition and machine translation techniques have been used to transform educational videos into easy-to-understand, organized content. The platform supports multilingual content and can transcribe, translate, summarize, and illustrate concepts in an effective manner. It also calculates the readability score of the extracted documents to ensure learners' understanding. The platform uses live data from YouTube, is 93% accurate, and responds quickly to search queries, in less than a second. The effective management of large data is handled by the four tier Command Query Responsibility Segregation (CQRS) architecture. Using their API simplifies the link up to YouTube and Google translate. This innovative approach provides solutions towards earing language barrier, saves learners time by helping them discover their needs quickly, and simplifies understanding of difficult subjects.

Keywords. Educational videos, NLP, automatic speech recognition, machine translation, transformers, large language models, semantic seasoning, machine learning.

1 Introduction

Online learning videos have changed the way people learn. Today, there are sites like YouTube that can store quadrillions of videos from basic math to even complex scientific information.

Although the huge store of videos is helpful, it becomes an issue in some cases when learners are unable to find suitable, well-made, age-appropriate videos. Most sites give more preference to videos with lots of likes and views, but that is not always the best way to know if the video is informative and helpful [7].

Most learners spend their precious time browsing numerous videos for obvious, simple, and reliable content. Most of the time, the video is too sophisticated, too simple, or insufficient in good reasoning. It is hard to determine how good an educational video is, how challenging it is, or how appropriate it is for primary issues. It can render learning frustrating and inefficient for learners who wish to learn independently.

The other big problem is the language used in these videos. Educational videos usually come in only one language, so people who do not

understand that language cannot use them. Even if there are subtitles or transcripts, they might just give long explanations. Transcripts can be unclear, inaccurate, or poorly organized, which makes it hard for learners to understand the content. Also, traditional approaches to organizing educational video content are largely based on manual curation, simplistic tagging of keywords, and likelihood-based suggestions, but these have typically failed to address modern learning needs [8] [21]. They ignore the effects of teaching proficiency, conceptual richness, and language friendliness, and give recommendations that are mostly irrelevant or incorrectly paired. Even though metadata-driven systems do not assess teaching effectiveness, static difficulty labeling does not cater to the variability of learner abilities.

While hand-transcribed captions and subtitles are useful they are not economically feasible in most cases and are not readily available for non-native learners [21] [17]. Recommendation systems often prioritize works of engagement over actual learning thus promoting visually appealing content of low quality. This means learners waste precious time through filtering from irrelevant content to knowledge, hence there is a need for intelligent technologies to enhance visibility of content and learning attainment. The present work aims to tackle these challenges by developing an e-learning platform with the ability to find suitable content, and getting individual assistance. Overall, the system aims to make learning easier by providing clean and organized materials and AI-mandated enhancement in comprehension.

2 Literature Survey

Major improvements in the field of educational video content analysis led to the use of natural language processing (NLP) and deep learning.

There has been a great deal of progress from many connected research efforts over the years.

Thus, articulating the level of competence of these technologies made it easier to explore, organize and convey educational content in a more relevant and experienced manner through which a learner used to interact with educational videos.

Using Latent Dirichlet Allocation (LDA), a three-level scheme for English education resources, document, topic and keyword was proposed which allows to organize the data and its descriptors properly. Moreover, it allows important educational topics to be highlighted and benefits from the parallel features of LDA [7]. In this context, CNN works together to identify keyframes and summarize the video material in an educational setting. By using both location and time, the framework allows to make logical summations.

It improves both precision and recall for the identification of important segments in videos [8].

An approach is presented that blends information from the video, its audio and its words. The system ensures that the summaries are organized as learners see the information. It showed that learners performed well in understanding situations when using comprehension [21]. For an educational video retrieval framework, semantics are understood using deep learning and captions for alignment. The system uses Automatic Speech Recognition (ASR) and NLP to determine the main topics in the lectures. An improved ability to find answers and greater satisfaction with searches were found in semantic content searches [17]. In another study, reinforcement learning is used to help agents to improve the quality and variety of educational material summarized. How significant the content is, will impact how long the summary is. The new method achieves better results in both relevance and coverage compared to the traditional methods [24].

By using graphs, educational videos are modeled and their information from the transcript and visual parts are combined into one semantic graph. The graph helps to track concepts as they develop and saves time in summarizing them. It is proven through testing that the system creates abstractive summaries by using the results of ASR with transformer models. It cleans up rough transcripts and offers a summary suited to each situation. It helps readers understand and remember the topic more easily than when using extractive methods [20] [4]. Summing up, the experts add attention functions to the time-related layers of transformer models for video processing. Using certain events and descriptions

allows us to maintain a consistent version of the story. When education achieves its main aims, strong performance metrics are also present [22]. First, information from the audio and visual data is extracted using the Sense2Video pipeline; next, the Bidirectional Encoder Representations from Transformers (BERT) technology is applied for summarization.

Multimodal embeddings in artificial intelligence allow us to abstract information in much more detail. When looking at these texts side by side, it is easy to see that both the focus on important parts of the text and understanding of the text are dealt with very well [13]. Using Generative Pre-trained Transformer (GPT), the system looks at the transcribed lectures and summarizes them, as well as asking questions.

From answering these questions, learners discover new things and look at their achievements. The evaluation suggests that learning and mastering concepts is improved for groups that benefit from the tool [14]. To make educational videos accessible in other languages, the paper combines speech recognition method, language translation and summary generation.

The Large Language Models (LLM) allow the system to guarantee the information means the same thing in every tongue. Experiments that mix multiple languages are useful and make information more accessible. Since there is so much information, users can try to understand concepts without encountering success. Another study proposes a Transformer method to offer more meaningful search results. The Video to Text API uses GPT-3 to provide accurate transcription, closed captions and indexing of your videos. With this approach, people can more easily look for and use professional videos [11, 10]. Android based Mobile Learning Platform Using Latent Semantic Analysis (AML-LSA) resolves this problem by using Latent Semantic Analysis, running on an Android device and integrating an online database and speech-recognition technologies. Learners receive feedback on their work immediately after submitting it [26]. Using questionnaires to spot a person's traditional learning style is both time-consuming and not dynamic.

A research suggests a new The Felder-Silverman Learning-Style Model (FSLSM) technique based on deep learning and topic modeling that can be used dynamically. By studying interaction data on the system through fuzzy logic and LDA techniques, it decides on the learning styles of each learner. Thanks to the model, detection is more accurate, and learning material is adapted to each learner [12]. Lecture videos are difficult to use because they follow one continuous sequence; the summaries from AI are meant to fix this problem. A summary is built by analyzing the images, their layout and the kind of font used during the ad [19]. In this analysis, reviews of MOOC courses are examined using Python web crawlers and LDA topic modeling to improve MOOC video design for aesthetic education. Someone learning a language underlines the importance of focusing on teaching abilities, the lessons, the level of exposure and video production. So, it is advisable to use effective methods like stories, interactivity and a pleasing design in your videos [5].

When educational videos are long, summing them up helps; this study focuses on LDA to efficiently separate the main ideas from the subtitles. It adds subtitles, generates a short summary from keywords and includes additional content which prolongs the episodes. It can be seen from the results that LDA gives better precision and recall scores than TF-IDF and LSA, however, it results in vague topics when it encounters mixed-topic documents [1]. The use of Collapsed Gibbs Sampling in LDA is studied to see how it helps assign topics to words effectively [15]. To overcome the shortcomings of language and domain in sentiment analysis, this paper develops Latent Semantic Scaling that connects seed words and word embeddings to separate documents along several axes using a semi-supervised approach. In combination with English and Japanese corporate uses, it provides outcomes like Lexicoder and suggests a different strategy to enhance the vector aspects [25]. In this case, the lexicon models of TextBlob help analyze the comments people post on YouTube education videos. Most of the emotions expressed were neutral, with positive and negative emotions the least seen. LDA reports that animation,

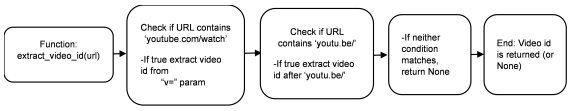

Fig. 1. Flow diagram of fetching video details

background music and clarity in the video's messages impact how viewers respond to it [6]. It points out that audio/video resources from educators are less popular in the archives due to the lack of smart ways to conduct searches. It includes a system where you can look up sentences and search for them by looking at letters or timing [2].

Another work introduced Probabilistic Latent Semantic Analysis (LSA) and outlined some of the main challenges it faces when being computed. In the beginning LDA method was used, but now, adopting NLP and distributions, it can operate more broadly. Neural networks are aided by the priority given, Fisher kernels and the training of machines. It includes a list of important areas that require further study [9].It produces subtitles using speech recognition and uses NLP to summarize the text it extracts from videos. It combines different techniques to achieve enhanced results. Reviews indicate that NLP/ML in summaries ensures main points are covered, thereby saving time for users watching video [3] [16].

The researchers use NLP techniques to analyze many video transcripts, captions, comments and metadata. Text summarization, sentiment analysis, topic modeling and deep learning are among the methods reviewed in this field. The discussion covers obstacles and future changes in NLP that may enhance how online videos are streamed [23].

3 Methodology

This project employs a multistage approach for effective processing and analysis of educational video content in a variety of languages and categories, with applied use of Automatic Speech Recognition (ASR), Language Detection or Machine Translation, followed by advanced transduction based on Large Language Model (LLM) based NLP reasoning. The layered pipeline that came with this enabled accurate transcription and translation of videos are also semantically understood and presented to the learners in learner friendly format. The aim of this system is to fill the gap between raw, unstructured video data and more accessible structured answers that can be used to educate individuals.

3.1 Steps of the Methodology

The steps of the methodology are described below.

3.1.1 Extract Video ID

A foundational step of the system involved the extraction of the video ID out of a YouTube URL. `extract_video_id` is intended as a function that would parse YouTube URLs of any kind (e.g. `https://www.youtube.com/watch?v=videoID` or `https://youtu.be/videoID`). It recognizes the part of the URL that corresponds to the unique video ID, that is, a string of characters which can uniquely identify the video across YouTube's platform, from which we determine its appropriate API URL. However, the system is able to parse the URL and extract this ID from the URL (refer Figure 1), which is used by it in subsequent API calls to retrieve more video specific data.

3.1.2 Fetch Video Metadata

Our system initially fetches the unique video ID by calling to the YouTube's Data API v3.

It is a video ID to be used as a primary reference to get detailed metadata regarding a particular video. The information is returned from YouTube's API in a JSON format that can be both human readable as well as machine interpretable.

Usually, this includes the video's title, description, publication date, tags and the name of the channel, video's view count, likes, and comments.

These first constitute the first step in the contextual understanding or knowledge about the system provided, this gives it high-level information about the purpose and subject of the video, and its popularity. Title and description are critical for the system identifying the overall theme the content may be about. It is usually the title that is taking

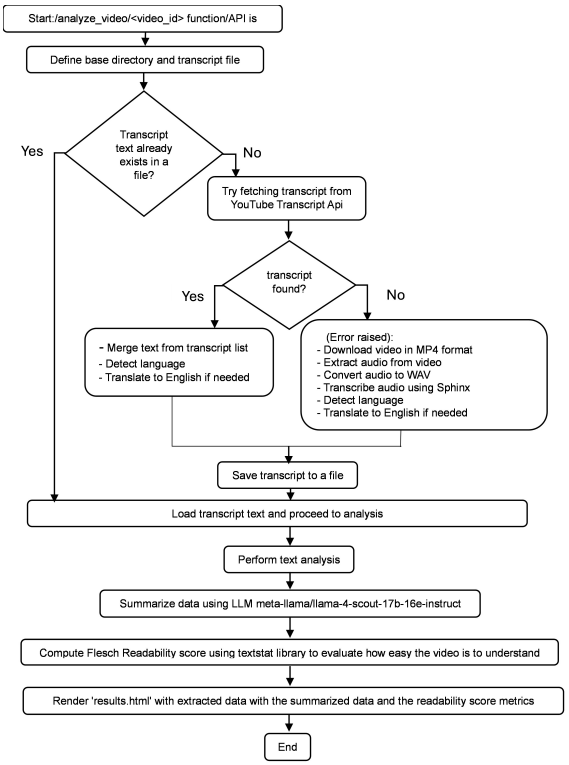

Fig. 2. Flow diagram for video summarization and readability scoring

the form of the topic of interest in a smaller way, and there is the description which will give an extra narrative or the contextual details that could provide a hint of how deep, technical nature, and educational intent of the video is.

For example, if the video is about science, the description may refer to certain theories, processes, or terms, and the system can then be used for tasks that come later after transcript segmentation, example mapping, etc., to determine the target audience of the video.

3.1.3 Retrieve Transcript

One of the most important parts of processing YouTube content is retrieval of the Transcript. The transcript is available via YouTube's own Transcript API if the video has been manually captioned or has auto generated by YouTube. First, if the system finds a locally stored transcript, it can save time not having to repeat the redundant fetches.

In case of no local transcript, the system will try to ask YouTube's API (`YouTubeTranscriptAPI`) to get the transcript in the languages available.

This transcript is then fetched and if it is not in English, the system uses the Google Translate API to translate it to English so that it can be compatible with the further analysis tools. It translates the video such that if the video is in a low resource/foreign language, the system can still process and analyze the content. This is a key step in making sure the transcript can be used in the subsequent NLP and text analysis steps.

3.1.4 Fallback to Audio-Based Transcription

When there is no transcript or when the provided transcript is not usable (e.g. it is incomplete, barely readable, etc.), or in simpler words, if no transcript is provided, the system goes to the audio-based transcription. Using `yt_dlp`, a flexible tool that supports video downloading in different formats, the video is initially downloaded. The video file is then converted to audio and is saved as an audio file. The spoken content is extracted from the video file and is provided to the system using `MoviePy` which will remove the audio track in the file and just leave the content. Finally, this is converted to the `WAV` format (16-bit PCM at 16kHz) which is generally compatible with most speech recognition systems. Then the system splits the audio into pieces that can be processed, around 30 seconds long to make the transcription more exact.

Then, real time transcription occurs with either the CMU Sphinx offline model or Google Speech Recognition. This fallback process guarantees the system will be able to transcribe spoken words even in instances when it does not have access to a transcript so further analysis can continue.

3.1.5 NLP Preprocessing

Once the transcript is obtained, the system performs NLP preprocessing using the spaCy library on it. Tokenization and removal of stop words and punctuations are performed in this stage.

3.1.6 Text Analysis and Readability

In this stage the system measures word count, sentence length and lexical diversity to estimate the length of the video, complexity of the idea and variety of language used respectively. The system then calculates the readability score using Flesch-Kincaid readability index [18].

3.1.7 Knowledge & Summary Generation

The system can gather relevant information, definitions, explanations, and important facts to support viewers in getting more from the video. Here, it involves recognizing jargon and terms particular to that field that should be simplified for most audiences.

We achieve this goal by running the `meta-llama/llama-4-scout-17b-16e-instruct` model accessed through Groq Cloud on their developer plan, as it is both effective and with affordable interface cost at $1 per million tokens.

The deployment has been done using Groq Cloud public API. With this model, we can speedily and precisely obtain necessary knowledge and generate summaries where response time ranges from 5–10 ms per token latency. For factual accuracy the model depends on pre-trained data on books, research papers, web content etc.

The model adopts the biases of the training data e.g. regional, demographic, cultural, topic specific, content preference patters etc. Following the first processing, the system uses its own custom features to find and compress the shortest and most meaningful sentences, themes and topics in the transcript. Furthermore, the system shows graphs and collections of words to demonstrate the presence and connection of terms in the transcript.

This matters a lot because the video's main things are boiled down into a user-friendly format, and through interesting and useful graphics. Gaining important knowledge from the video becomes more straightforward for users. Refer to Figure 2 for the flow diagram for retrieving summarized data.

3.2 Backend Using 4 Layer Command Query Responsibility Segregation (CQRS)

3.2.1 Controller Layer

To process a YouTube video in the Controller Layer, it receives the URL in the request and forwards it to the Business Layer. It also handles responses and errors.

3.2.2 Business Layer

The Business Layer implements the CQRS pattern and separates commands and queries, which improves scalability and system performance.

3.2.3 Data Access Layer

The Data Access Layer abstracts how the data is stored, retrieved, and manipulated through repositories and ORM tools. It serves as an interface between the Business Layer and the databases.

3.2.4 Database Layer

The Database Layer stores entities such as user profiles, video metadata, transcripts, analysis results, and summaries. This layer helps maintain data integrity and ensures efficient performance.

4 Experimental Setup

4.1 Software Tools, Libraries & Frameworks Used

For building the backend of the system, Python web framework, *Flask*, was used since it is simple, lightweight and it's very easy to handle requests and connect to other services. It allowed for quick and easy development of API's and playing nicely with other tools. The *Google API Client* was employed so that the system can interface with *YouTube Data API* to fetch video information.

With the YouTube Transcript API, it was easy to get subtitles and captions of YouTube videos. The *Requests* library was used because of its support for simple web requests to get things like video transcripts or metadata from the internet. The transcripts were collected, and their language was

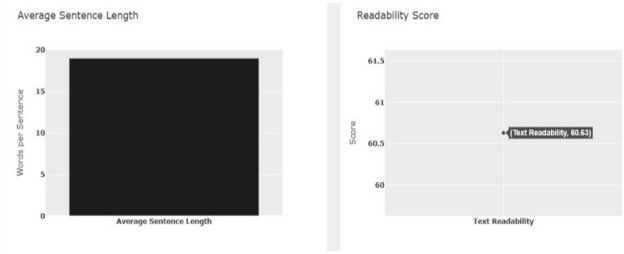

Fig. 3 (a) Fig. 3 (b)

Fig. 3. Average Sentence Length and Readability Score for a video

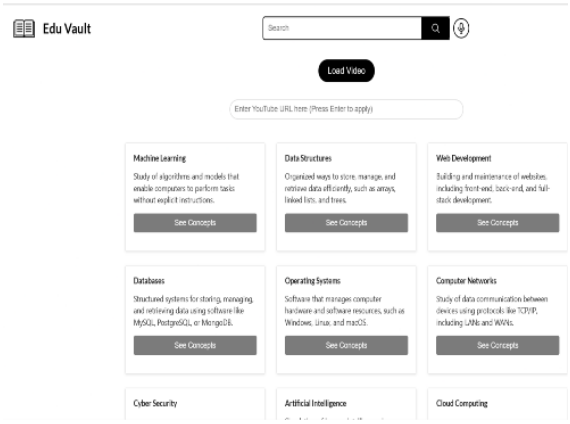

Fig. 4. Home Page

checked using the Python library *LangDetect*. This knowledge helped the system to know whether a translation to English was needed. The audio of the video was taken from the video using *yt-dlp* which downloaded it. In this work, the current audio was converted to text with the help of tools like *Google Speech API* using the *SpeechRecognition* library of Python.

The backbone of the Web API was built using *C#* as the major programming language. In the application, there were two key libraries used for database operations. *MediaTR* offered ORM (Object Relational Mapping) capabilities that simplified mails access in the database through easy, effective and rapid data communication, thereby making the process simplified. Alongside that, *Dapper* was also a lightweight ORM tool to run fast SQL queries and map database results to *C#* objects in a fast and easy way. These technologies came together to create an optimal pipeline between the application and its database.

MySQL provided an efficient and reliable backend for storing user profiles, notes, and processed educational material for user profile. And *HTML*, *CSS*, *JavaScript* are used for making a responsive and interactive front-end interface. *AJAX* and *Fetch API* allowed to do asynchronous communication with the backend.

4.2 Complexity Assessment Using Graphs

The system generates advanced NLP and ML based visualizations for each of the educational videos, which are the Relevance Graph and Complexity Graph, which serve to increase the learner's experience. First the model extracts the semantic features from the transcript and compares them to the user intended topic. To extract the concepts of the video and the learner's objective, as part of an alignment component, techniques such as concept extraction, keyword clustering and semantic similarity scoring are used to measure how closely the video content matches the learner's goal. It allows the system to produce a dynamic systemic context relevant video score.

Figure 3 (a) and (b) together represent visual outputs that calculate linguistic complexity and the readability of the educational content. Figure 3 (a) consists of a bar chart, in which the bar represents the average sentence length in terms of the number of words per sentence, on the y-axis. Figure 3 (b) then shows normalized readability score with a scatter plot. This visualization gives insight into how easy the text is to understand, from the point of view of language and structure of the sentence.

5 Results and Discussion

Once users successfully sign up or login using *Edu Vault* login page, they are redirected to the *Edu Vault* home page (refer to Figure 4). This interface gives us a way through many input methods to explore educational content. On top, there is a standard search bar on the top with facilities for *text-based search* as well as a feature for *audio-based search* using voice input. A *YouTube*

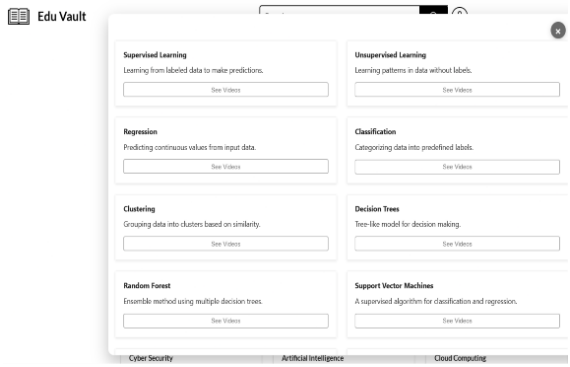

Fig. 5. Subtopics

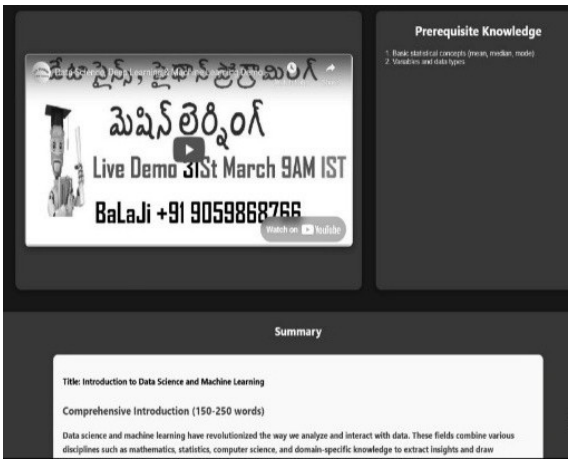

Fig. 6. Automatically generated prerequisite concepts and structured summary for a Telugu educational video

URL input field below lets the users directly paste the links that will carry with them the specific video-based content retrieval and analysis.

The main part of the interface displays the list of educational domains like *Machine Learning*, *Data Structures*, *Web Development*, *Databases*, *Operating Systems*, *Computer Networks* etc. The short informative description of each topic is then followed by a navigation option *"See Concepts"* which bring the user to the explained concepts in detail.

In this layout space is available to learn interactively and the system provides access to different technical subjects in a friendly environment. A short explanatory text is written under each subtopic that indicates the main concept of the topic and makes it clear for users what is the central idea of the subtopic. Below each description there is a *"See Videos"* button that allows users to access curated video content that is directly related to that subtopic (refer to Figure 5).

5.1 Multilingual Speech Transcription

5.2 Video Resource Display for Selected Subtopics

Figure 12 depicts how the educational video interface is shown to the users after selecting a particular subtopic. Each video is embedded under the interface directly with *title*, *embedded YouTube player*, *duration* easily viewable, and a *"Explore"* button. On the other hand, the *Explore* feature is provided for users to get more details about the video or its metadata, for instance, key topics covered, timestamps, or summary content.

Through this layout, it allows the easy organization of multiple videos pertaining to the subtopic that a learner would choose to allow them to engage with educational content based on both interest and time availability. The *Prerequisite Knowledge Panel* also points out the foundational concepts in videos (e.g. arrays, modulo operations and Big O notation etc) needed to understand the video content. With this, learners are also able to self-assess whether they have the necessary background before proceeding.

The *Summary Panel*, located below the video section, comes with an independent summary of the content. In this one it will have a structured summary, with an *introduction*, *core explanation*, *key concepts*, a *tutorial-style implementation*, and *conclusion*.

5.3 Saved Notes Interface for Educational Videos

In Figure 13, we can see the *"My Saved Notes"* interface where the users can record and bring back their personalized notes attached to specific educational videos. Users can save key insights, summaries and explanations in natural language for each video that they

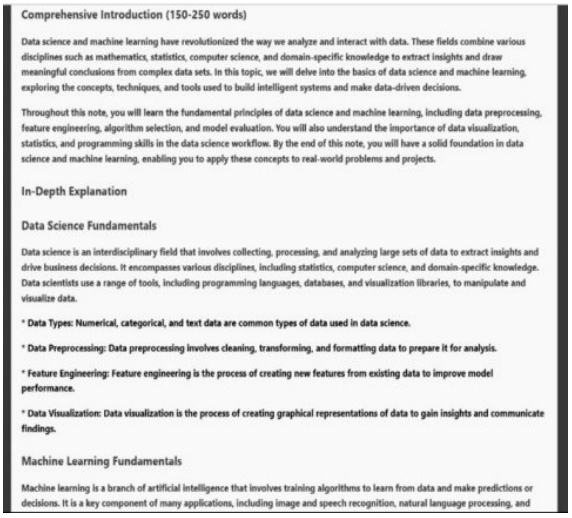

Fig. 7. System output illustrating prerequisite knowledge extraction and English summary generation from a Telugu video

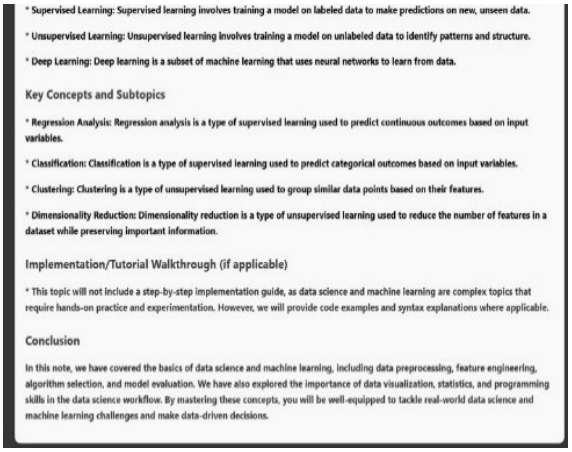

Fig. 8. Example of prerequisite knowledge identification and AI-based summary creation for a video originally in Telugu

have viewed before using the platform. Below the video thumbnails correspond to these notes presented in a structured manner. If the *"See Note"* button is clicked, we bring up a modal window with the full saved note's content. This helps learners memorize the right points from educational videos and come back again for revision or advanced understanding.

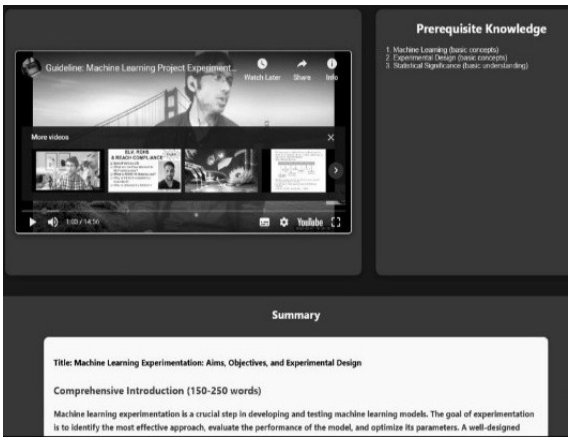

Fig. 9. Automatically generated prerequisite concepts and structured summary for a Bengali educational video

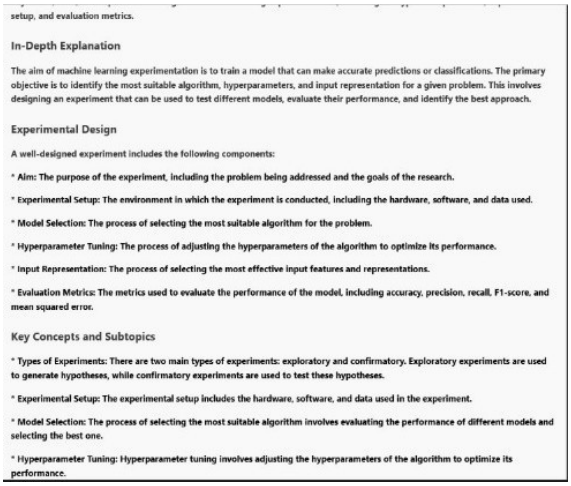

Fig. 10. System output illustrating prerequisite knowledge extraction and English summary generation from a Bengali video

5.4 Integrated AI Chatbot for Educational Query Assistance

Figure 14 shows an AI-powered chatbot interface in the *Edu Vault* platform. It allows users to directly ask academic questions within the learning environment. The chatbot, on the other hand, responds with contextual educational-related responses to a wide spectrum of topics, examples, as it is shown with *"What is Hashing in DSA"*.

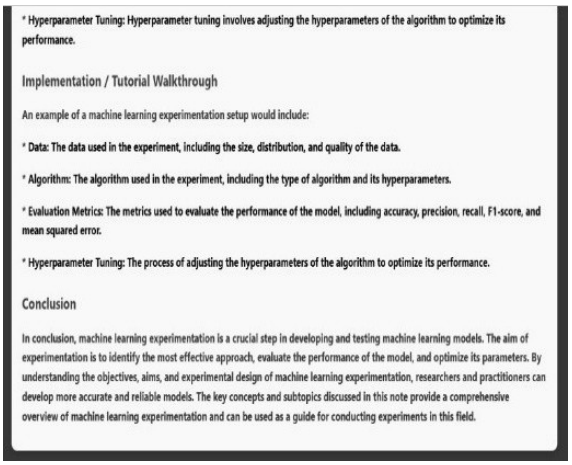

Fig. 11. Example of prerequisite knowledge identification and AI-based summary creation for a video originally in Bengali

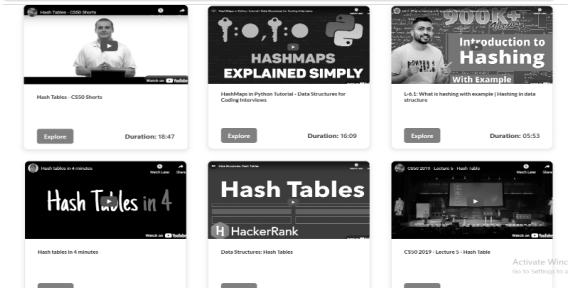

Fig. 12. Videos on a Topic

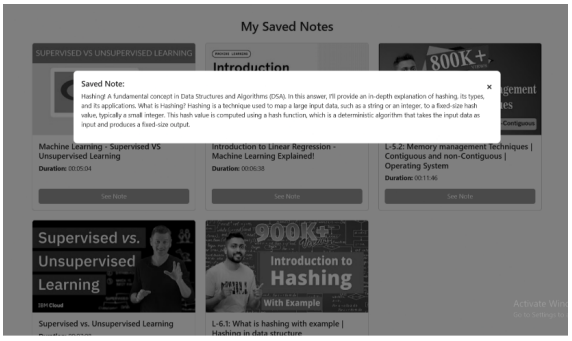

Fig. 13. Video Notes

The chatbot will add value by providing instant clarification about a concept when it is needed, without the user needing to leave the existing context, the page. It can support such structured explanations as *definitions*, *use cases* (e.g., indexing, verification, data compression), and formatted responses containing *highlighted keywords*.

The in-page assistant helps one reduce time spent trying to seek external resources and it effectively enhances idea reinforcement and, subsequently, the level of engagement and efficiency of learning.

5.5 Performance Evaluation

A semantic similarity approach based on *paraphrase-MiniLM-L6-v2* of *Sentence Transformers* was used to evaluate the performance of the proposed system. We use cosine similarity to find out if a user's search query is similar to a video's title and description.

Accuracy, precision, recall and F1-score are calculated to measure how well our system retrieves information (refer to Figure 15).

These demonstrate that the system can provide relevant information for topics in STEM (Science, Technology, Engineering and Mathematics). For a specific query ("Red Black Trees"), the system returned 50 videos with API response time 0.99 seconds and with the following confusion matrix values derived from system predictions and manual relevance labeling by human experts:

— **True Positives (TP):** 46, Videos correctly identified as relevant.

— **False Positives (FP):** 2, Videos incorrectly identified as relevant.

— **True Negatives (TN):** 0, Videos correctly identified as irrelevant.

— **False Negatives (FN):** 2, Relevant videos missed by the system.

The reason for the misclassification of the 4 videos out of 50, roots from the model's automatic generation of summaries, where the key term for search may have been omitted.

From this confusion matrix, standard information retrieval metrics are computed as follows:

— **Accuracy:** Proportion of total correct predictions 92%

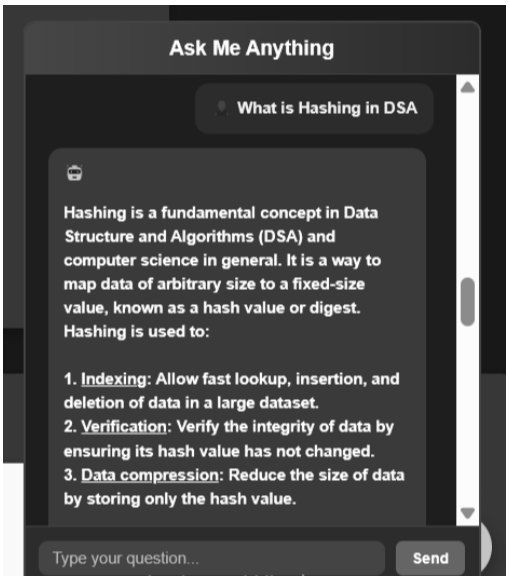

Fig. 14. Chatbot

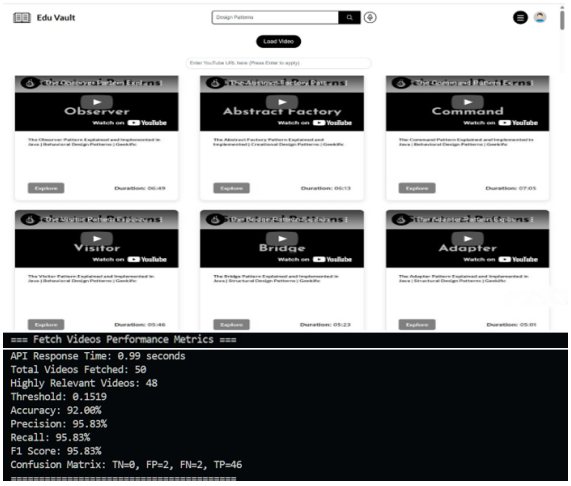

Fig. 15. Systems Performance

— **Precision:** Proportion of predicted relevant videos that are truly relevant 95.83%

— **Recall:** Proportion of truly relevant videos that are retrieved 95.83%

— **F1-Score:** Harmonic mean of precision and recall 95.83% (0.958).

Unlike other systems, the proposed system pulls up-to-date video content from the YouTube Data API. Because of this, the data keeps changing.

All systems use labeled and interpretable data, but our model operates on noisy and continuous data, and thus it has been discovered to be flexible and adaptable. And found that the accuracy of the present system is consistent in different scenarios.

To ensure our system's excellence, we reviewed different articles from the literature for performance metrics and compared them with our live-system metrics. Our model is found to be highly precise and has high recall as well as high accuracy in generating data, demonstrating scalability and contextual comprehension despite noisy, real-time metadata. A comparison of performance metrics with the existing works is shown in Table 1.

Moreover, the present work, based on the `meta-llama/llama-4-scout-17b-16e-instruct` model, outperforms other recent transformer-based multimodal systems, such as CLIP-based video summarizers, Gemini, or Gemini 1.5 Flash, mainly in speed and low token cost, which makes the system more accessible to a vast community of learners across the globe.

User feedback has been collected through a survey on subject matter experts and groups of target learners to measure the effectiveness of the system and overall users' satisfaction while using the system. The feedback analysis shows high user satisfaction towards accuracy of summaries, readability scores, or chatbot responses and significant reduction in search time in retrieving appropriate learning videos to meet users' educational needs.

6 Conclusion

The present work developed an intelligent e-learning platform, making educational videos easier to use, organized and more helpful to learners around the world. We used AI tools, including ASR, language detection, machine translation and NLP, to change videos speech into properly organized and understandable learning content. It supports multilingual content, allowing learners from diverse language backgrounds to benefit equally. Our system is unique as it

Table 1. Comparison of performance metrics with existing methods

Method	Accuracy	Precision	Recall	F1-Score	Data Type
Our System (Live)	93%	93%	88%	0.906	Dynamic, real-time
Educational Videos Subtitles' Summarization Using LDA [1]	N/A	71.66%	86.66%	0.7814	Static
Text Summarization using NLP [4]	N/A	88%	58%	0.68	Static

can transcribe, translate, summarize, pick out the most important details, measure the difficulty and illustrate concepts and generates the prerequisite knowledge which is required to watch the content.

And the model generates the summary for the video the user likes to explore in an easy way and intuitive way. The processed text undergoes NLP tasks like cleaning, tokenization, stop-word removal, and lemmatization, followed by readability scoring to ensure learner understanding. LLM models are then applied to generate summaries, extract key points, effectively reducing load. It also provides a feature of saving notes for the video that the user is interested in storing for future use or revision. Generally, other systems cannot handle dynamic content or need human input, whereas our platform uses live data from YouTube. Because it is 93% accurate on average and responds in less than a second to generate the content based on the search query, it is considered very reliable.

There is also a simple and useful dashboard, and chatbot help and personal learning achievement tracking. The system uses the latest Transformer and Gemini models to thoroughly analyze what is written, offering summaries that are easy for anyone to understand. By applying AI to modern ML, we have made the platform innovative and reliable. Therefore, the present approach is better because it saves learners time, helps them discover what they need, and eases their understanding of even the most difficult subjects.

References

1. **Alrumiah, S. S., Al-Shargabi, A. A. (2022).** Educational videos subtitles' summarization using latent dirichlet allocation and length enhancement. Computers, Materials & Continua, Vol. 70, No. 3.
2. **Arazzi, M., Ferretti, M., Nocera, A. (2023).** Semantic hierarchical indexing for online video lessons using natural language processing. Department of Electrical, Computer and Biomedical Engineering, University of Pavia.
3. **Aswin, V. B., et al. (2021).** Nlp-driven ensemble-based automatic subtitle generation and semantic video summarization technique. In Advances in Artificial Intelligence and Data Engineering: Select Proceedings of AIDE 2019. Springer Singapore.
4. **Balaji, N., Kumari, D., Bhavatarini, N., Megha, N., Kumar, S. (2022).** Text summarization using nlp technique. 2022 International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), IEEE, pp. 30–35.
5. **Cao, J. (2024).** Intelligence empowers aesthetic education: A study of mooc video production based on lda topic modelling. Proceedings of the 2024 5th International Conference on Education, Knowledge and Information Management (ICEKIM), Shanghai, China.
6. **Chalkias, I., Tzafilkou, K., Karapiperis, D., Tjortjis, C. (2023).** Learning analytics on youtube educational videos: Exploring sentiment analysis methods and topic clustering. Electronics.
7. **Du, W., Zhu, H., Saeheaw, T. (2021).** Application of the lda model to semantic annotation of web-based english educational

resources. Journal of Web Engineering, Vol. 20, No. 4, pp. 1113–1136.

8. **Fei, X., Saravanan, V. (2020).** An lda based model for semantic annotation of web english educational resources. Journal of Intelligent & Fuzzy Systems.
9. **Figuera, P., García Bringas, P. (2024).** Revisiting probabilistic latent semantic analysis: Extensions, challenges and insights. Faculty of Engineering, University of Deusto, Vol. 12, No. 1.
10. **Hanumesh, M., Sankar, K. B., Anitha, G., Pattanaik, B. (2024).** Semantic enrichment of video content using nlp transformer networks. Available online, March 31, 2024.
11. **He, Z., et al. (2022).** Unsupervised learning style classification for learning path generation in online education platforms. Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining.
12. **Hussain, T., Yu, L., Asim, M., Ahmed, A., Wani, M. A. (2024).** Enhancing e-learning adaptability with automated learning style identification and sentiment analysis: A hybrid deep learning approach for smart education. IEEE Transactions on Learning Technologies, Vol. 17, pp. 1–15.
13. **K., H. K., Palakurthi, A. K., Putnala, V., Kumar K., A. (2020).** Smart college chatbot using ml and python. 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India.
14. **Kumar, A., et al. (2019).** Chatbot in python. International Research Journal of Engineering and Technology (IRJET), Vol. 6, No. 11.
15. **Ogunwale, Y. E., Ajinaja, M. O. (2023).** Application research on semantic analysis using latent dirichlet allocation and collapsed gibbs sampling for topic discovery. Asian Journal of Research in Computer Science, Vol. 16, No. 4, pp. 445–452.
16. **Panthagani, V. B., et al. (2024).** Youtube transcript summarizer. 2024 5th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), IEEE.
17. **Patil, R., Buchade, A., Yadav, G., Sharma, N., Joshi, S., Bhokare, A. (2024).** Youtube video summarizer using asr. 2024 IEEE International Conference on Blockchain and Distributed Systems Security (ICBDS), Pune, India.
18. **Patterson, K., Street, J. A., Myachykov, A. (2024).** Phrasal frequency and literacy as predictors of individual differences in on-line processing and comprehension of english complex np subject-verb agreement. Journal of Cultural Cognitive Science, Vol. 8, pp. 247–274.
19. **Rahman, M. R., Koka, R. S., Shah, S. K., Solorio, T., Subhlok, J. (2024).** Enhancing lecture video navigation with ai generated summaries. Education and Information Technologies (EAIT).
20. **Rahul, Adhikari, S., Monika (2020).** Nlp based machine learning approaches for text summarization. 2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC), Erode, India.
21. **S, S. R., M, V., N, S., K, S. (2025).** Natural language processing-driven learning platform for personalized education. 2025 International Conference on Visual Analytics and Data Visualization (ICVADV), Tirunelveli, India, pp. 174–180. DOI: 10.1109/ICVADV63329.2025.10961687.
22. **Samandarov, E. (2025).** The architecture of educational platform based on machine learning. JMMCS, Vol. 124, No. 4, pp. 86–102.
23. **Singh, Y., et al. (2023).** Youtube video summarizer using nlp: A review. International Journal of Performability Engineering, Vol. 19, No. 12, pp. 817.
24. **Vayadande, K., Nemade, M., Parbhanikar, S., Rathod, S., Raut, A., Thorat, R. (2023).** Efficient content exploration on youtube: Automatic speech recognition-based video summarization. 2023 7th International Conference on Electronics, Communication and

Aerospace Technology (ICECA), Coimbatore, India.

25. **Watanabe, K. (2020).** Latent semantic scaling: A semisupervised text analysis technique for new domains and languages. Communication Methods and Measures, Vol. 15, No. 2, pp. 81–102.

26. **Zhang, D. (2024).** Application of android based mobile learning platform using latent semantic analysis. 2024 International Conference on Integrated Intelligence and Communication Systems (ICIICS), IEEE.

Article received on 30/05/2025; accepted on 27/07/2025.
**Corresponding author is Lopa Mandal.*