eVISION: Educational Video Impact Scoring through Multimodal Analysis and Machine Learning

MCS

Author:

Gul Raze

(Registration No: **00000431941**)

Supervisor

Lt Col Dr Yasir Awais Butt

A thesis submitted to the faculty of Computer Software Engineering Department, Military College of Signals, National University of Sciences and Technology, Islamabad, Pakistan in partial fulfilment of the requirements for the degree of Master of Science in Software Engineering

(Sep 2024)

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis written by Mr **Gul Raze, 00000431941**, of **Military College of Signals** has been vetted by undersigned, found complete in all respect as per NUST Statutes/Regulations, is free of plagiarism, errors and mistakes and is accepted as partial, fulfillment for award of MS/MPhil degree. It is further certified that necessary amendments as pointed out by GEC members of the student have been also incorporated in the said thesis.

Signature: ____________________

Name of Supervisor: Lt Col Dr. Yasir Awais Butt

Date: ____________________

Signature (HoD): ____________________ Brig Head of Dept of CSE Military College of Sigs (NUST)

Date: ____________________

Signature (Dean/Principal): ____________________

Date: 23/9/24

Brig Dean, MCS (NUST) (Asif Masood, Phd)

Form TH-4

NATIONAL UNIVERSITY OF SCIENCES & TECHNOLOGY

MASTER THESIS WORK

We hereby recommend that the dissertation prepared under our supervision by **Maj Gul Raze,** Regn No **00000431941** Titled: **"eVISION: Educational Video Impact Scoring through Multimodal Analysis and Machine Learning"** be accepted in partial fulfillment of the requirements for the award of **MS Software Engineering** degree.

Examination Committee Members

1. Name: Assoc Prof Dr. Asad Ullah Signature: ____________

2. Name: Assoc Prof Dr. Ihtesham Ul Islam Signature: ____________

Co-Supervisor's Name Brig Adnan Ahmed Khan, PhD Signature: ____________

Supervisor's Name: Lt Col Yasir Awais Butt, PhD Signature: ____________

Date: 19/9/24

Brig
Head of Dept of CSE
Mil College of Sigs (NUST)

Head of Department

19/9/24
Date

COUNTERSIGNED

Brig
Dean, MCS (NUST)
(Asif Masood, Phd)

Dean

Date: 23/9/24

CERTIFICATE OF APPROVAL

This is to certify that the research work presented in this thesis, entitled "**eVISION Educational Video Impact Scoring through Multimodal analysis and Machine Learning** was conducted by Mr. **Gul Raze** under the supervision of **Lt Col Dr. Yasir Awais Butt**. No part of this thesis has been submitted anywhere else for any other degree. This thesis is submitted to the **Department of Computer Software Engineering** in partial fulfillment of the requirements for the degree of Master of Science in Field of Computer **Software Engineering Department of Military College of Signals, National University of Sciences and Technology, Islamabad**.

Student Name: **Gul Raze** Signature:

Examination Committee:

a) External Examiner 1: **Assoc Prof Dr Asad ullah** Signature:
(Department of Computer Software Engineering)

b) External Examiner 2: **Assoc Prof Dr Ihtisham ul Islam** Signature:
(Department of Computer Software Engineering)

Name of Supervisor: **Lt Col Dr. Yasir Awais Butt** Signature:

Brig
Head of Dept of CSE
Mil College of Sigs (NUST)

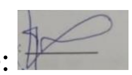
Name of Dean/HOD: **Brig Adnan Ahmed Khan, PhD** Signature:

<u>PLAGIARISM UNDERTAKING</u>

I solemnly declare that research work presented in the thesis titled "**eVISION: Educational Video Impact Scoring through Multimodal Analysis and Machine Learning"** is solely my research work with no significant contribution from any other person. Small contribution/ help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero-tolerance policy of the HEC and National University of Sciences and Technology (NUST), Islamabad towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred/cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS degree, the University reserves the rights to withdraw/revoke my MS degree and that HEC and NUST, Islamabad has the right to publish my name on the HEC/University website on which names of students are placed who submitted plagiarized thesis.

Student Signature: 

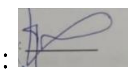
Name: **<u>Gul Raze</u>**

Date: Sep 2024

AUTHOR'S DECLARATION

I **Gul Raze** hereby state that my MS thesis titled "**eVISION: Educational Video Impact Scoring through Multimodal Analysis and Machine Learning**" is my own work and has not been submitted previously by me for taking any degree from National University of Sciences and Technology, Islamabad or anywhere else in the country/ world.

At any time if my statement is found to be incorrect even after I graduate, the university has the right to withdraw my MS degree.

Student Signature: 

Name: **Gul Raze**

Date: Sep 2024

<u>DEDICATION</u>

“In the name of Allah, the most Beneficent, the most Merciful"

I dedicate this thesis to my family, friends, and teachers who supported me each step of the way, especially my parents.

ACKNOWLEDGMENTS

All praises to Allah for the strengths and His blessing in completing this thesis.

I am thankful to Allah Almighty for His unwavering support and the countless blessings He has bestowed upon me throughout my life. It is by His grace that I have achieved this milestone. Indeed, this research work could not have been completed without His support.

I am extremely grateful to my thesis supervisor, **Lt Col Dr Yasir Awais Butt**, and Co-Supervisor, **Brig Adnan Ahmed Khan, PhD** for their unwavering support and guidance throughout my thesis. Their knowledge, expertise, and dedication have been a source of inspiration to me, and I am grateful for the time and effort they have invested in my success. Whenever I encountered any difficulties, they were always available to offer their assistance and provide me with insightful feedback.

I also extend my gratitude to my GEC members, **Assoc Prof Dr Asad Ihsan** and **Assoc Prof Dr Ihtesham ul Islam**, for their continuous availability and support throughout my degree, both in the coursework and thesis. Their knowledge and expertise have significantly contributed to the accomplishment of my task.

ABSTRACT

In the modern learning landscape, educational videos and Massive Open Online Courses (MOOCs) have become a central tool for delivering content, offering a vibrant and an effective way to engage learners. These videos offer an interactive medium for disseminating educational content, meeting the needs of various learning styles and preferences. However, as the availability and accessibility of educational videos on digital educational platforms continue to expand, ensuring their quality and effectiveness becomes a critical challenge for the students, educators and instructional workforce. Currently, there is no standardized criterion for evaluating and rating of educational videos. This lack of assessment can lead to inconsistent quality in the educational material, making the evaluation process time consuming and potentially undermining the learning experience. This paper proposes a novel data-driven framework for evaluating and scoring educational videos based on their metadata (number of views, likes/ dislikes, comments sentiment etc) and the content of their transcripts (spoken content within the video). The goal is to create a framework that enables automated evaluation of video content, providing learners, educators, and content creators with a more comprehensive understanding to enhance the learning experience. To support this analysis, specialized datasets are created for both metadata and transcript, which focuses on these important factors of educational videos. A scoring mechanism has been devised through user feedback supported by statistical techniques for establishing a baseline for educational videos. Machine learning regression-based models and deep learning models are used to predict the scores, with their accuracy checked using Mean Squared Error (MSE), Mean Absolute Error (MAE) and R squared. The XGBoost model emerged as the most effective model for metadata-based predictions, while Support Vector Regressor (SVR) excelled in transcript-based data. After predicting scores for metadata and transcript for each video, an overall score is determined by averaging these values. This overall score serves as a reliable indicator of the educational value of the video, considering both its popularity and the strength of its content. These results highlight the potential of machine learning models to effectively predict and rate educational video quality, offering a robust framework for a more objective assessment of digital educational content and contributing to the enhancement of global education standards.

Table of Contents

List of Figures

List of Tables

List of Abbreviation

ML	--------------------	Machine Learning
MOOC	--------------------	Massive Open Online Course
HLRBM	--------------------	Hybrid Logistic Regression Bagging Model
SVM	--------------------	Support Vector Machine
NB	--------------------	Naïve Bayes
RF	--------------------	Random Forest
KNN	--------------------	K-Nearest Neighbor
DT	--------------------	Decision Tree
LR	--------------------	Logistic Regression
ANN	--------------------	Artificial Neural Network
AB	--------------------	AdaBoost
SMOTE	--------------------	Synthetic Minority Oversampling Technique
GA	--------------------	Genetic Algorithm
MSE	--------------------	Mean Squared Error
MAE	--------------------	Mean Absolute Error
TF	--------------------	Term Frequency
IDF	--------------------	Inverse Document Frequency

Chapter 1 - Introduction

1.1 Overview

It is quite clear that in the modern world, the education has become the most important factor in defining the future [1]. But there remains a significant challenge as access to quality education to all is still an elusive goal. There are still constraints towards educational access such as lack of funds, distance to the institutes, and poor facilities. These barriers only serve to sustain disparities in the acquisition of knowledge and skills.

Digital education is a great way to help with many tough problems. It uses technology to give people access to learning tools. This is especially helpful for learners who live far away or have trouble getting around. Online courses, fun educational videos, and virtual classes can connect students with schools they might not be able to reach [2]. Educational videos are really popular, they're flexible and make learning fun [3]. But even with all these benefits, checking how good these videos are can be tricky. We usually look at things like teacher ratings or what students say. But those things can be pretty subjective. So, they don't always tell us the full story about how effective a video really is.

This research aims to address this challenge by developing a standardized framework for evaluating the quality of educational videos. By analyzing both metadata and transcript content, the study seeks to create a reliable and objective scoring system. A specialized dataset has been constructed, and various machine learning and deep learning algorithms were employed to predict educational video scores based on their features. The results demonstrate the effectiveness of machine learning models in accurately assessing video quality, offering a valuable tool for educators about educational resources. The findings of this research contribute to developing more objective and reliable assessment methods for educational videos. This can improve online learning experiences and help bridge the education gap.

1.2 Importance of Education

It is a basic human right and a key factor in driving personal and societal development. Education equips individuals with the knowledge and skills necessary for meaningful engagement in their community's economic, social, and political spheres. It is essential for reducing poverty, improving health outcomes, and promoting gender equality. Despite the recognized importance of education, a significant number of children and adolescents around the world are deprived of basic educational opportunities. According to a report by UNESCO, approximately 258 million children and youth between the ages of 6 and18 were out of school globally in 2022 [4]. The lack of education is not evenly distributed across the globe. Certain regions, particularly in sub-Saharan Africa and South Asia, have the highest numbers of out-of-school children. Additionally, many students who are enrolled in school do not achieve basic proficiency levels in reading, exacerbating the global education crisis [5].

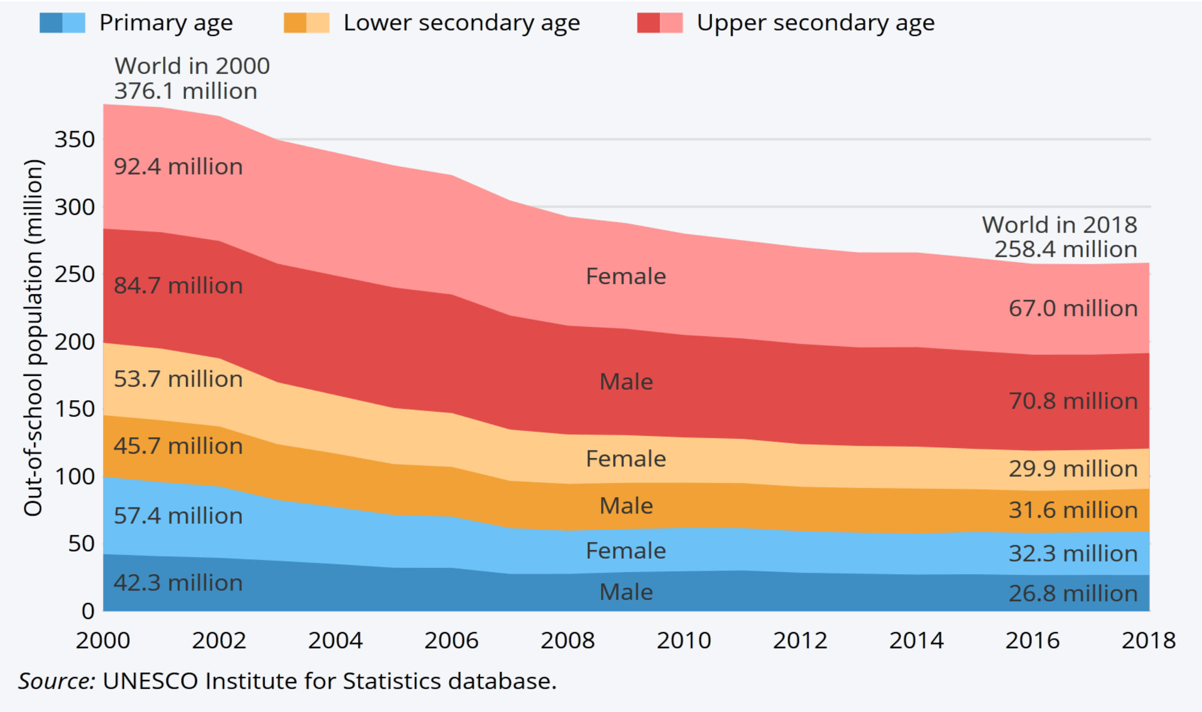


Figure 1.1: Global number of Out-of-School Children 2000-2018

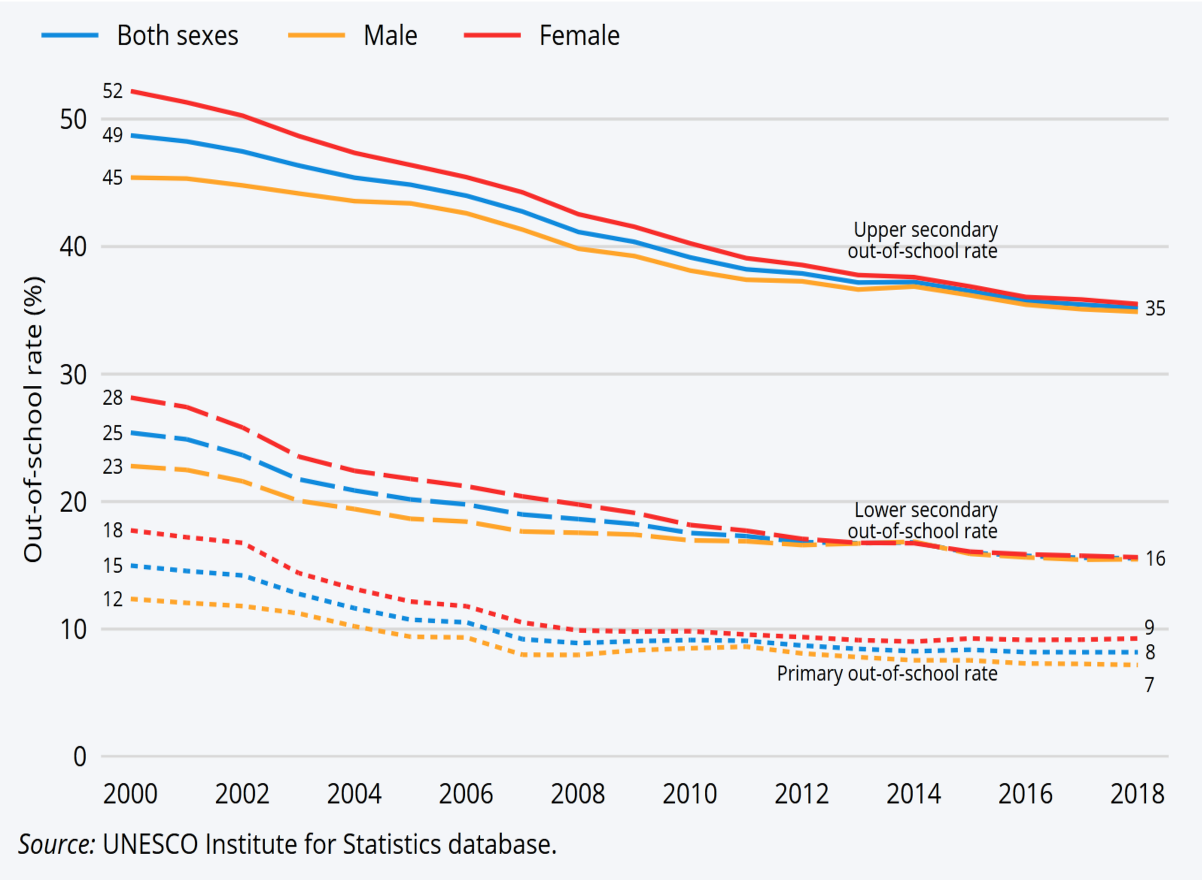


Figure 1.2: Global out of School rate by Age Group and Sex 2000-2018

1.3 Implications of Lack of Education

Lack of education has serious consequences for individuals and societies:

- **Economic Impact** A lack of education leads to lower income potential, perpetuating cycles of poverty. Countries with lower education rates often experience slower economic growth.
- **Health Outcomes** Education is closely linked to health. Educated individuals are more likely to understand health information, leading to better health outcomes for themselves and their families. Educated women, in particular, are more likely to seek healthcare services and have healthier children.
- **Social Inequality** Education is a powerful tool for social mobility. Without access to education, lowered groups, including girls, ethnic minorities, and people with disabilities, are more likely to remain trapped in cycles of poverty and inequality.

1.4 Digital Education

It refers to the use of digital technologies and online platforms to deliver educational content and facilitate learning. Digital education has emerged as an essential component of modern learning. Digital education can be delivered through various platforms, such as:

- **Learning management systems (LMS):** Online platforms that provide tools for delivering, managing, and tracking educational content.
- **Mobile apps**: Educational apps that can be accessed on smartphones and tablets.
- **Social media**: Platforms like Facebook, Twitter, and YouTube can be used to deliver educational content and facilitate discussions. Mobile apps: Educational apps that can be accessed on smartphones and tablets.

1.5 Role of Digital Education in Transforming Access to Education

It offers students the ability to access educational materials anytime and anywhere. Educational videos have become increasingly dominant in modern learning environment, offering a dynamic and engaging medium for delivering content to users enhancing the learning experience.

Figure 1.3: Online Digital Educational Platforms

One of the most significant advantages of digital education is the availability of free resources. Various platforms, such as Khan Academy, Udemy, and YouTube provide quality educational materials at no cost, helping students and educators access a wealth of knowledge without financial strain. These platforms offer a diverse array of resources, including interactive lessons, video tutorials, and practice exercises, allowing learners to engage deeply with various subjects. Traditionally, the evaluation of educational videos has often relied on subjective assessments or simplistic metrics such as view counts or completion rates. While these measures provide some insight into user engagement, but they fail to capture the detailed aspects of learning outcomes and the effectiveness of instructional content. Moreover, with the increase of online learning platforms and massive open online courses (MOOCs), the sheer volume of available educational videos makes manual evaluation impractical and inefficient.

Previous research on educational video efficacy has encountered several key challenges that hinder comprehensive assessment and analysis. One major issue is the reliance on single modality analysis, such as text-based sentiment analysis or visual content analysis. While these approaches have provided valuable insights, they often fail to capture the holistic nature of educational videos and may overlook important cues present in other modalities.

1.6 Problem Statement

With the growing reliance on digital educational platforms, it is essential to have a reliable and objective method for evaluating the quality of educational videos and their impact. However, current practices in assessing educational videos lack standardization and often fail to provide a comprehensive evaluation of content relevance and audience engagement. This research aims to address this gap by developing a framework for a robust, data-driven multimodal approach for a more objective assessment of educational videos, thereby contributing to improved global education standards and supporting better decision-making by educators, students, and content creators.

1.7 Research Objectives:

- **Analyzing Multimodal Data** Examine the metadata and textual components of educational videos to identify patterns, trends, and correlations that contribute to their effectiveness in facilitating learning.

- **Understanding User Engagement** Examine factors such as viewer attention, interaction patterns, and emotional responses to assess the user engagement.
- **Development of Tool to Assess Educational Videos** Primary goal of this research is to develop a tool which can assess the efficacy of educational videos in enhancing learning outcomes by investigating their impact on students' knowledge acquisition, comprehension and retention.

1.8 Research Motivation

The motivation behind this research stems from a practical industry need identified by alnafi.com, a platform offering a range of online courses, diplomas, and certifications at affordable prices. As the platform sought to enhance its offerings, it became apparent that there was a need for a reliable tool to evaluate the effectiveness of its educational videos. The challenge was to develop a method for assessing educational video quality based on a variety of data points, ensuring that students receive high-quality educational content. This research aims to address this need by creating a robust framework for evaluating educational videos, thus providing valuable insights into their effectiveness and helping platforms like alnafi.com improve the quality of their educational resources.

1.9 Key Contributions

The key contributions of this research are: -

- The development of structured datasets for metadata and transcript analysis includes meticulous data processing, normalization, and feature engineering, making the data well-suited for machine learning applications.
- Collection of educational videos from diverse online platforms, ensuring comprehensive coverage of content from various sources.

- Employing APIs or web scraping techniques to retrieve features of educational videos which involves navigating legal, ethical, and technical challenges, such as API limitations, rate restrictions, and website terms of service.
- Converted video content into audio files which is needed to create transcripts, by using dependable tools and methods ensuring accuracy of video-to-audio conversion processes.
- The use of speech recognition technologies, like the Whisper API, to extract text from audio files ensures high-quality transcripts, which are essential for accurate and reliable analysis.
- To accurately interpret educational content, implemented advanced linguistic techniques in transcript analysis to extract features in the textual content, which involves a deep understanding of natural language processing (NLP) and text analysis.
- The use of speech recognition technologies, like the Whisper API, to extract text from audio files ensures high-quality transcripts, which are essential for accurate and reliable textual analysis.
- Developed a scoring system for educational videos after collection of student feedback through survey forms for chunk of videos.
- Proposed a statistical method for manually providing a score to educational videos as provision of score through survey forms for large number of videos was not possible.
- Application of different machine learning algorithms to predict scores for educational videos, using both metadata and transcript data, this research has significantly enhanced the accuracy and effectiveness of evaluating video quality and relevance.
- Combining scores from both metadata and textual component analyses, and averaging them to provide an overall score to educational videos, enhanced the

evaluation process by offering a comprehensive and balanced assessment of video quality.

1.10 Areas of Application

- **Education** Learning practices will be influenced by this research which will ultimately improve their design and delivery of instructional materials, curriculum development, and teaching strategies.
- **E-Learning Platforms** Online learning platforms can utilize this research to improve the effectiveness of their educational video content, leading to better engagement and retention among learners.
- **Corporate Training** Companies and organizations often use educational videos for employee training and development. This research can help optimize training materials to ensure maximum effectiveness, leading to better skill acquisition and performance improvement among employees.
- **Content Creation and Production** The findings of this research can empower content creators and producers in the media and entertainment industry to develop educational videos that are both engaging and effective.

1.11 Relevance to National Needs

This research is highly relevant for our country. It focuses on making education better & helping create a skilled workforce. It's also part of the Higher Education Commission's (HEC) effort to offer Massive Open Online Courses (MOOCs) to all universities. This show's a big need for a way check how well these courses work. By improving education at every level and closing gaps in learning, especially in far-off or underserved areas, this study can help everyone have a fair shot at quality education. Plus, it tackles some tough problems like having not enough resources or too many kids in one classroom. It promotes video-based learning that is easy to scale and doesn't cost too much. This can provide useful ideas for those making decisions and teachers too.

1.12 Paper Organization

This thesis is divided into seven chapters:

1. **Chapter 1**: This chapter includes the basic introduction, problem statement, research objectives, motivation, and contribution.
2. **Chapter 2**: This chapter provides an overview of relevant literature, encompassing articles pertinent to the scope of this study.
3. **Chapter 3**: This chapter presents the material and methods.
4. **Chapter 4**: This chapter delivers the experimental results.
5. **Chapter 5**: This chapter describes the discussion.
6. **Chapter 6**: This chapter concludes the report and highlights the direction for future work.

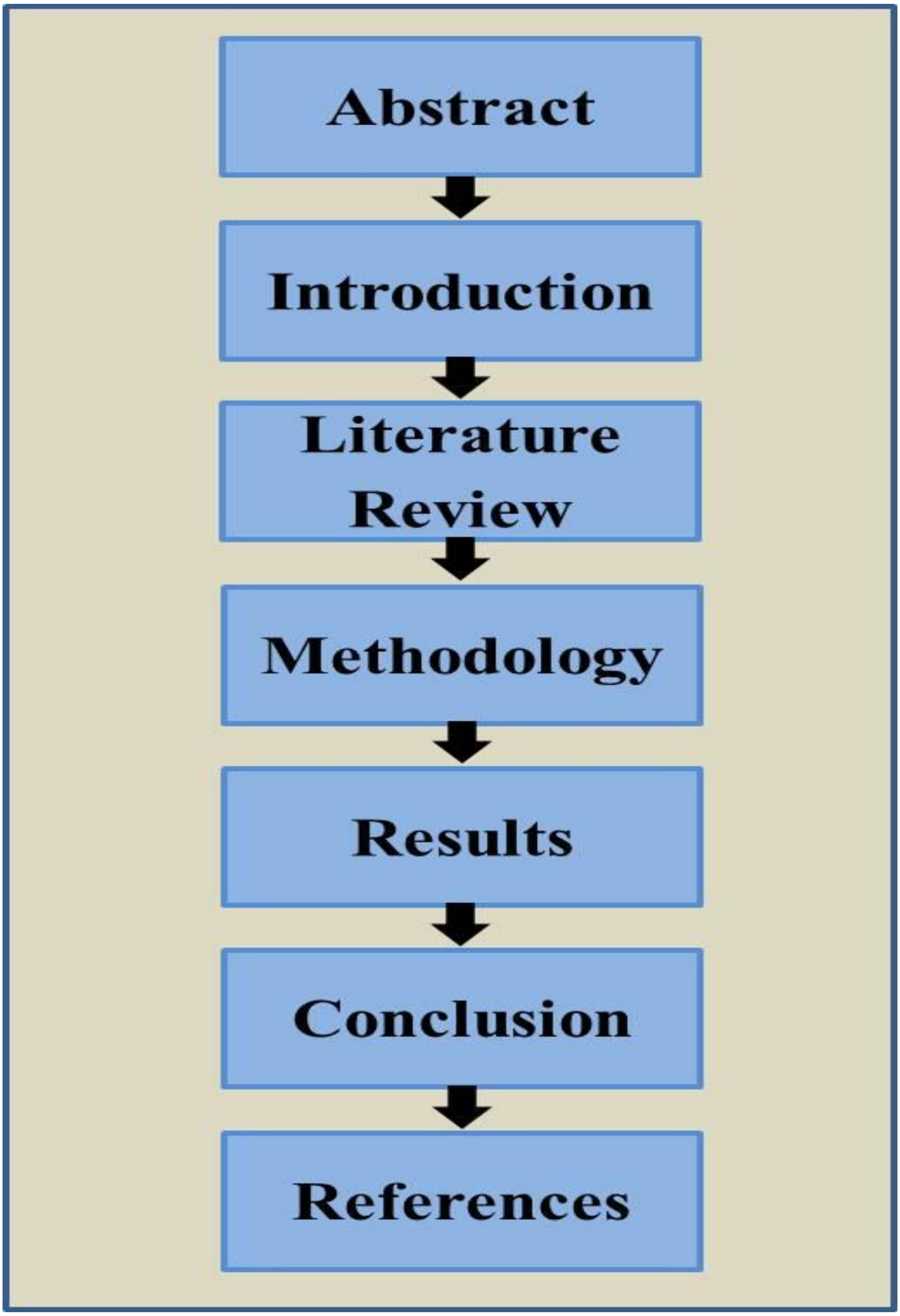


Figure 1.4: Taxonomy of Thesis

Chapter 2 - Literature Review

2.1 Overview

In Chapter 2, the literature review and related work concerning the evaluation of educational videos using machine learning techniques are discussed. This section focuses on the prior research conducted by various scholars and the outcomes of their studies. With an increase demand for online learning platforms, there is a growing interest in developing tools that can accurately assess the quality of educational videos. The main reasons for this include the need to improve content delivery and ensure effective learning experiences and create standardized evaluation criteria for educational resources. The objective of this section is to review the existing research on assessing video quality, focusing on three key areas: techniques for educational video evaluation, the role of metadata and transcript analysis in video assessment, and the application of machine learning algorithms in this context.

2.2 Related Work on Educational Videos Metadata

In [6] a novel methodology is proposed for classifying YouTube videos as educational or non-educational using advanced machine learning models, particularly Convolutional Neural Networks (CNNs), to enhance the accuracy of video classification. It expands the dataset to include 500 features derived from user comments, thereby improving the classification process. Additionally, the research aims to enhance the LOIS (Learning Objects Intelligent Search) recommendation system to provide more accurate recommendations of high-quality educational videos, incorporating sentiment analysis to evaluate user feedback. This work suggests integrating sentiment analysis into the recommendation system to evaluate the quality of videos based on user feedback, thereby recommending videos that are not only educational but also positively received by viewers. Overall, this research seeks to streamline the process of locating educational content on YouTube, making it more efficient for both teachers and students using user significant number of comments (approximately 738,500) extracted from 500 educational videos. Experimental results showed that a Convolutional

Neural Network was able to differentiate educational videos from non-educational ones with an accuracy rate of 95.71%. In [7] the authors proposed a methodology that uses combination of classification, association, and clustering techniques to analyze the factors influencing video trends. This involves data collection, preprocessing, feature extraction, and the application of various machine learning algorithms to analyze the data and identify patterns related to trending videos. The dataset consists of YouTube video metadata, which includes attributes such as video titles, descriptions, tags, and engagement metrics (likes, dislikes, comments, views). The paper applies various machine learning algorithms, including classification algorithms (such as Decision Trees, Random Forests, or Support Vector Machines), association rule learning techniques, and clustering algorithms (like K-Means or Hierarchical Clustering). The proposed model achieves an accuracy of 95%, indicating its effectiveness in classifying and predicting trending videos based on the analyzed features.

Another study [8] carried out by Riyan Amanda et al (2020) proposes to classify YouTube data into categories such as Arts and Science using machine learning techniques. The dataset consists of links, titles, descriptions, and search queries scraped from YouTube. The methodology follows an experimental approach, including data collection, preprocessing, model proposal, testing, and evaluation. The algorithms compared in the study are Random Forest, Support Vector Machine (SVM), and Naïve Bayes. The results indicate that the Naïve Bayes model achieved the highest accuracy of 88%, while the Random Forest and SVM models both achieved an accuracy of 82%. Similarly in [9] written by Michael Carmichael et al (2023) highlighted how educational videos significantly increase student engagement and improve learning experience. This study employs surveys and experimental methods to analyze the video content in higher education settings. The findings suggest that integrating videos into teaching strategies can lead to more interactive and effective learning experiences. Another study [10] by Rhitabrat Pokharel et al (2021) proposes a method to classify YouTube comments into sentiment categories (positive, negative, neutral) and sentence types (questions, statements). The methodology includes data collection through web scraping, preprocessing to clean the data, and feature extraction using TF-IDF and word embeddings. The study employs machine learning models like SVM, Random Forest, and Naïve Bayes, achieving an accuracy of approximately 85% for sentiment classification and 80% for sentence type classification. This classification helps content creators better understand viewer feedback and improve engagement.

In [11] proposes a method to classify YouTube videos based on their quality using seven different machine learning algorithms. The study compares the performance of K-Nearest Neighbors (KNN), Decision Trees, Naïve Bayes, Random Forest, Logistic Regression, ExtraTrees, and Stochastic Gradient Descent (SGD) classifiers. The Random Forest algorithm achieved the highest accuracy, making it the most effective model for this classification task. The findings suggest that machine learning can effectively classify video quality, aiding content creators in improving their video production standards. Similarly in [12, 13,14] researchers investigated the quality and reliability of YouTube videos on spondylolisthesis, rotator cuff and ovarian cysts using Global Quality Score (GQS) criteria. All these studies analyzed 50 videos, finding that videos uploaded by healthcare professionals had higher reliability scores but lower view counts. The results highlight the need for more accurate and reliable health information on YouTube to better serve public needs. In [15] authors proposed a model to analyze sentiments in comments on YouTube educational videos using machine learning and deep learning models. The study conducted experiments on imbalanced and balanced datasets, achieving the best accuracy of 96% with SVC, RF, and DL models using oversampling and SMOTE techniques. This research introduces a novel dataset based on Arabic YouTube educational videos, contributing significantly to data mining and related research areas. Moreover, Abdulhadi Shoufan et al (2019) author of [16] investigates how well educational videos on YouTube support cognitive features based on the cognitive theory of multimedia learning. The study uses learning analytics to analyze various cognitive aspects of educational videos and their impact on student learning. The methodology involves evaluating video content against cognitive principles and assessing their effectiveness in enhancing learning outcomes. The study employs regression analysis, achieving an adjusted R-square value of 68%, indicating that further research is needed to identify additional cognitive features in educational videos.

2.3 Related Work on Transcript/ Textual Analysis

Mohsen Mesgar et al (2018) author of [17] proposes a model that captures the semantic flow between adjacent sentences to assess text quality. The methodology involves representing sentence semantics with vectors and using a convolutional neural network (CNN) to encode coherence patterns. The model was evaluated on readability assessment and essay scoring tasks, achieving state-of-the-art results in readability assessment and significantly improving essay scoring performance, with an accuracy rate of 85% for readability assessment. In another study [18] authors introduced an automated essay grading system that leverages Natural Language Processing (NLP) techniques to evaluate and score essays. The system is designed to mimic the manual grading process by analyzing various linguistic and cognitive features of the text. Transformer-based language models such as BERT and RoBERTa used in this paper to handle long-term dependencies and extract meaningful insights from the essays, even when they are poorly written. The system demonstrates high accuracy and efficiency, making it a valuable tool for educational institutions. The paper reports that BERT achieved the highest accuracy in grading essays, demonstrating its effectiveness in understanding and evaluating complex linguistic features. Moreover, in [19] researchers conducted the analysis of transcripts in two stages, in first stage it compares existing automated text quality assessment programs, identifying their strengths and limitations. In the second stage, the authors develop additional criteria specifically tailored to the assessment of technical translation texts. Using expert evaluations, these new criteria are ranked and integrated into the assessment framework. The findings highlight the importance of specialized criteria in accurately evaluating the quality of educational content in distance learning environments, particularly for technical subjects.

In the paper [20], authored by Muddassira Arshad, Muhammad Murtaza Yousaf, and Syed Mansoor Sarwar, investigates the readability of scientific learning resources, particularly in the domains of Computer Science (CS), Machine Learning (ML), Software Engineering (SE), and Natural Language Processing (NLP). The authors introduce the AGREE dataset, comprising 42,850 learning resources, including research papers, lecture notes, and Wikipedia content. The study employs 14 readability indices and 12 lexical measures to assess text readability. The Extra Tree classifier was found to perform best on the AGREE dataset, demonstrating high accuracy and efficiency. The research highlights the lack of consensus among readability measures for shorter texts but notes improved accuracy for longer texts. The findings contribute significantly to the field by providing datasets and readability measures that can be used to train deep learning models, develop recommender systems, and assist in curriculum planning within the CS domain. In [21] paper presents two approaches for assessing the quality of software requirements using Natural Language Processing (NLP) techniques. The first approach is based on static syntactic and semantic analysis, which examines the syntactic structure and semantic properties of sentences to identify potential ambiguities and vagueness. The second approach relies on requirement guidelines, which include rules to avoid imprecise language and passive verb forms. By automating the quality assessment process, these methods aim to improve the clarity and comprehensibility of requirements, thereby enhancing the overall quality of the software development process. The study highlights the importance of well-written requirements in reducing errors and improving the efficiency of automatic information extraction in electronic design automation.

2.4 Research Gap Analysis

Prior research is mostly focused on single modalities, like text or metadata by itself, without thoroughly examining the synergistic impacts of metadata and textual characteristics. In order to close this gap, this study looks at the interactions and effects that textual analysis (readability and content analysis) and metadata (like counts, view counts, and comment metrics) have on the overall efficacy of instructional video content. Through the integration of these components, the research hopes to reveal fresh insights that are often missed in the literature today, thereby providing a more thorough knowledge of how various parts of instructional videos contribute to learner engagement and educational results.

2.5 Summary

The exploration of metadata in evaluating digital content reveals its increasing role in understanding user engagement and the overall impact of educational videos. Metrics such as view counts, likes, and comment sentiments provide valuable insights, but many studies focus only on surface-level interactions rather than deeply assessing content quality. This suggests a gap in how metadata can be standardized to effectively measure educational value. In contrast, transcript analysis has advanced significantly with the rise of natural language processing (NLP) tools. By examining the textual content of educational videos, researchers have uncovered ways to assess coherence, relevance, and educational worth using techniques like sentiment analysis and topic modeling. While transcript-based approaches offer deeper insights, integrating these findings with metadata analysis remains an underexplored area.
In sum, while considerable progress has been made in both metadata and transcript analysis, the next step is to combine these two approaches. Doing so could offer a more comprehensive and accurate evaluation of educational videos, bridging the gap between surface metrics and content quality.

Chapter 3 - Research Methodology

3.1 Overview

In Chapter 3, the systematic process and techniques employed to achieve the research objectives, focused on developing a multimodal machine learning framework for evaluating educational videos, are outlined. This section provides a detailed account of how data was collected, preprocessed, and analyzed, with an emphasis on both video metadata and textual components. The approach taken for model development and the prediction of educational scores is described, highlighting the thoughtful integration of multiple steps to ensure the research is comprehensive and reliable. Special emphasis is placed on improving the precision and efficiency of video content evaluation, ensuring that the methodology aligns with and effectively supports the core objectives of the research

3.2 Proposed eVISION Architecture

The proposed architecture of this research word is illustrated in Figure 3.1. The proposed framework for evaluating educational videos leverages two new specialized datasets which have been created for this research, Metadata and transcript dataset. Metadata dataset is produced by fetching data from YouTube through Google’s YouTube API which provided all the relevant details of educational videos like view count, like count etc. Feature engineering techniques were applied on the data to form new features to check the interaction and engagement of user with the educational video. Transcript dataset is generated by converting video into text using OpenAI’s Whisper API. Linguistic techniques, including coherence, keyword frequency and readability scores, are applied to extract valuable metrics from the transcripts. Both datasets are normalized using Min-max normalization to ensure data quality. Using combination of feedback survey and statistical techniques, score of the target variable is annotated creating a target variable for both modalities. Machine learning models such as XGboost, Linear Regression, Decision Tree are trained on these both datasets. Performance of these models are evaluated using Mean absolute error (MAE), Mean squared error (MSE) and R Squared to ensure precise predictions of educational video quality. This framework can serve as a benchmark for gauging the efficacy of high quality educational videos.

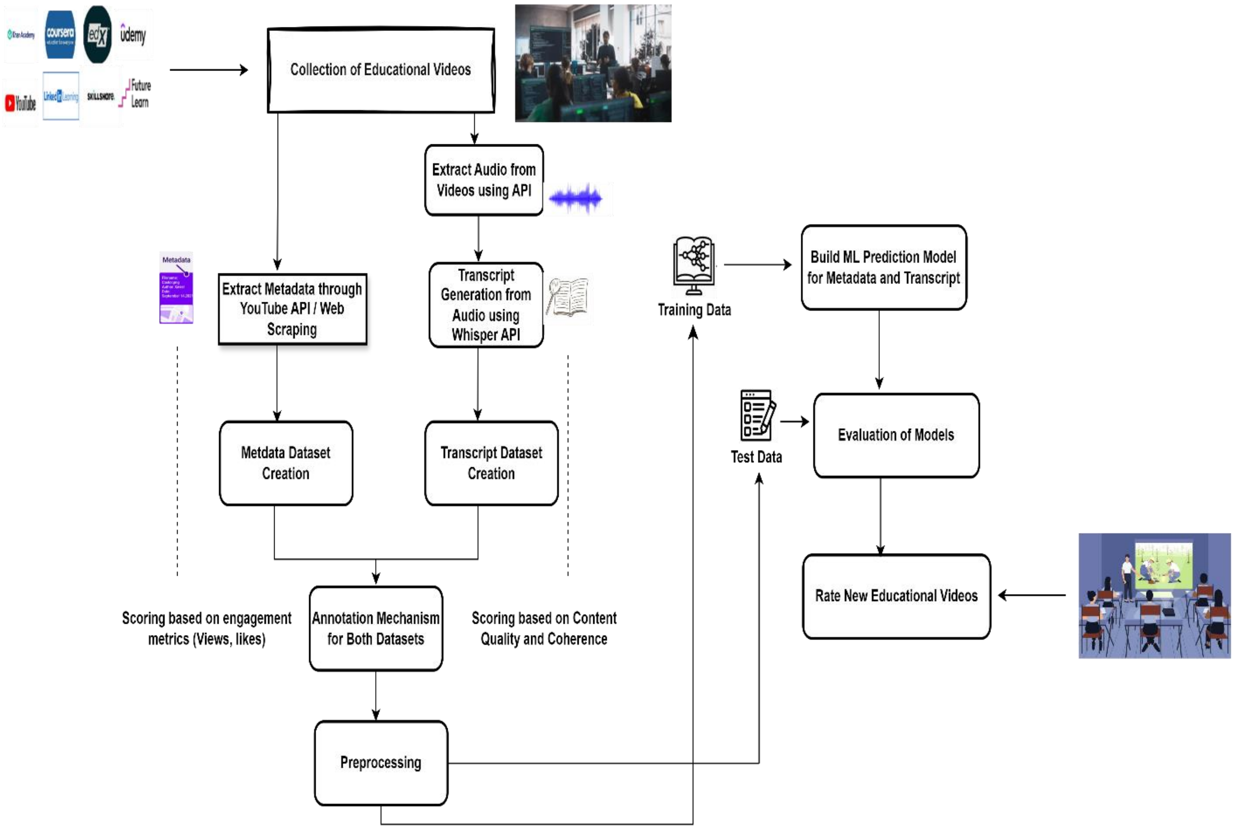


Figure 3.1: Proposed eVISION Architecture

3.3 Formulation of Metadata and Transcript Datasets

In order to develop a machine learning framework capable of assessing the quality and effectiveness of educational videos, two specialized datasets were meticulously created, providing a crucial base for the evaluation process. Both of these datasets contain features of 1,050 educational videos. The metadata dataset includes key user engagement metrics like view count, like count, comment count, and comment polarity. Meanwhile, the transcript dataset focuses on analyzing the readability, coherence, and topic relevance of the educational content. By combining these datasets, the research offers a well-rounded evaluation that integrates both engagement data and content quality insights.

3.4 Collection of Educational Videos

To ensure that the dataset is representative of diverse educational content, a specific set of criteria is used to select 1,050 educational videos. These criteria include:

- **Topics:** Videos covered a wide range of academic subjects, including (Computer Science, English, Technology, Engineering, Mathematics).
- **Duration**: Videos were selected across varying lengths, from short tutorials (under 10 minutes) to longer lectures (30-60 minutes) to capture different teaching formats.
- **Popularity:** Videos with varying engagement levels (including both popular and less popular ones) were selected to capture a broad spectrum of user interactions and preferences.
- **Source Credibility:** Only videos from trusted and well-known educational platforms were included to ensure the content was of high educational value.

3.5 Metadata Dataset

In order to assess the quality and effectiveness of educational videos, it was essential to collect comprehensive metadata that reflects both user engagement and video performance. Using the **YouTube Data API**, an efficient Python-based code was developed to automate the retrieval of these metadata features. YouTube API provides access to a variety of features and data points from YouTube videos and channels.

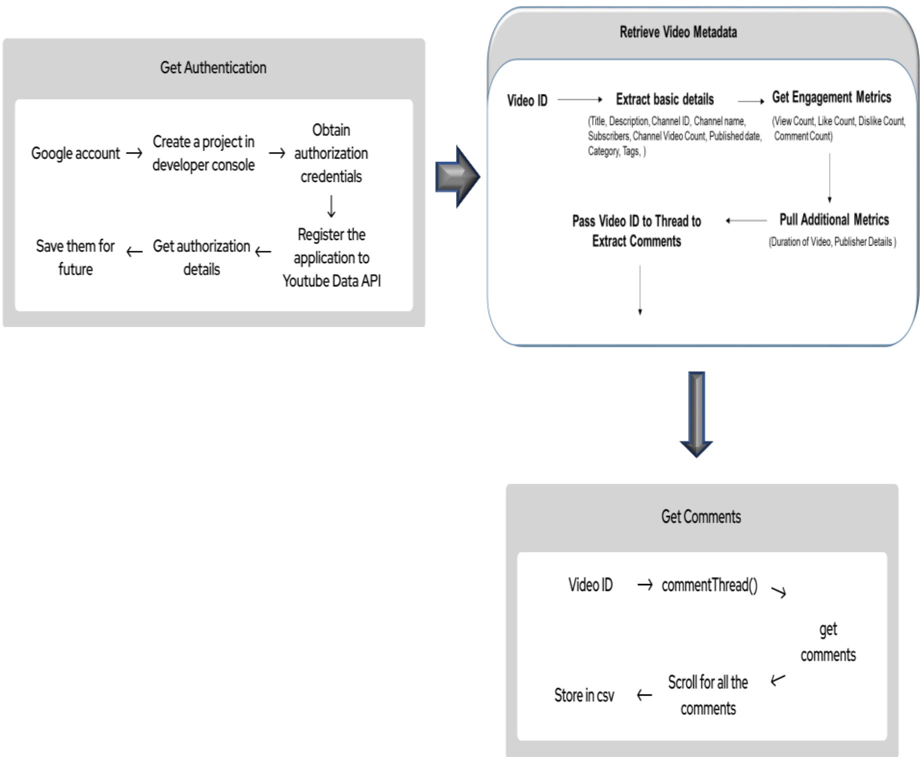


Figure 3.2: Process of Extracting Video Metadata Features using YouTube API

3.5.1 Features of Metadata Dataset

The extracted metadata features from YouTube videos provide comprehensive insights into various aspects of the content and its performance.

Table 3.1 Educational Video Metadata Features

Feature	**Description**	**Importance**
Video ID	A unique identifier assigned to each video by YouTube.	Essential for retrieving and distinguishing specific videos.
YouTube Link	The URL that directs to the video on YouTube.	Provides direct access to the video and facilitates sharing.
Channel Name	The name of the YouTube channel that uploaded the video.	Identifies the content creator and helps analyse channel performance.
Video Title	The title of the video as provided by the uploader.	Summarizes the video's content and aids in identification and search.
Video Description	A textual explanation of the video content provided by the uploader.	Offers additional context, enhances SEO, and informs viewers.
Published At	The date and time when the video was published on YouTube.	Helps analyse the video's age and publishing timing impact.
View Count	The total number of times the video has been viewed.	Indicates video popularity and reach.
Like Count	The number of positive reactions (likes) the video has received.	Reflects viewer approval and engagement.
Comment Count	The total number of comments posted on the video.	Shows the level of viewer interaction and engagement.

Subscriber Count	The total number of users subscribed to the channel.	Reflects the channel’s reach and potential audience size.
Channel Views	The total number of views across all videos on the channel.	Provides insight into the overall popularity and performance of the channel.
Total Videos	The total number of videos uploaded by the channel.	Indicates the channel’s content output and activity level.
Tag Count	The number of tags associated with the video.	Affects video discoverability and relevance.
Duration (Seconds)	The length of the video measured in seconds.	Impacts viewer engagement and retention.
Description Length	The number of characters in the video’s description.	Provides an idea of the depth and detail of the content summary.
Engineered Sentiment and Engagement Features		
Comments Average Polarity	The average sentiment polarity score of all comments of educational video.	Indicates how positively or negatively the audience reacted to the video.
Like to View Ratio	The ratio of likes to total views of the video.	Measures how engaging or likable the content is relative to its total views.
Comment to View Ratio	The ratio of comments to total views of the video.	Shows the level of interaction and engagement with the content.
Average Daily Views	The average number of views the video receives per day since its publication.	Provides insight into the video’s daily reach and popularity over time.
Interrogative Comment Count	The number of comments that contain a question or inquiry.	Reflects how much the video content sparks curiosity and interaction among viewers.
Engagement Rate	A combined score of user interactions (comments, likes) relative to views, reflecting user engagement.	A higher score indicates more active and involved viewers, signifying strong content engagement.

3.5.2 Videos Sentiment Score

To assess viewer sentiment for YouTube videos, comments were collected using the YouTube API and analyzed for sentiment polarity. Polarity refers to the measure of sentiment expressed in a piece of text. The process involved the following steps:

- **Collect Comments**: All comments for the YouTube videos were extracted using the YouTube API. This involves making API requests to gather comments and handling pagination if there are multiple pages of comments.
- **Preprocess Comments**: The text of each comment was preprocessed by converting it to lowercase, tokenizing, removing stopwords and punctuation, stemming, lemmatizing, and removing numeric values.
- **Analyze Sentiment:** A sentiment analysis tool, such as TextBlob, was used to assess the sentiment of each comment. TextBlob assigned a polarity score to each comment, ranging from -1 (very negative) to +1 (very positive), indicating the emotional tone of the text.
- **Calculate Average Polarity**: After computing the polarity score of each comment, average of all the scores was taken providing a summary of the emotional response expressed in the comments.

3.5.3 Feature Engineering of Metadata

To check engagement of users into the educational videos, new features were created using feature engineering: -

- **Average Daily Views**: As watchtime of any YouTube video is only restricted to the individual who owns the channel, so for this a new feature is created using the existing features by dividing by the total view count by the number of days since the video was published, determined by subtracting the published date from the data collection date.

  Average Daily Views = View Count / Days Since Published
- **Engagement Rate**: This feature is created to perceive how actively viewers are

interacting with the video. It was calculated by adding the total likes and comments, then dividing by the view count.

Engagement Rate = Likes + Comments / View Count

- **Like-to-View Ratio**: This metric measures the proportion of likes a video receives relative to its total view count. It indicates how much viewers appreciate the content relative to how many have watched it.

  Like-to-View Ratio = View Count / Like Count

- **Comments-to-View Ratio**: This feature measures the ratio of comments to the total number of views. It reflects the level of viewer engagement and interaction through comments.

  Comments-to-View Ratio = View Count / Comments Count

- **Interrogative Comment Count**: This metric counts the number of comments that contain questions. It provides insights into the level of viewer curiosity or need for clarification related to the video content.

3.6 Transcript Dataset

In the context of educational videos, a transcript is a spoken content within a video. The transcript captures every word spoken in the video, providing a complete and accurate text representation of the audio. This dataset and its associated metrics offer a comprehensive view of the spoken content quality in educational videos, providing valuable insights into how well the content engages and informs viewers.

3.6.1 Transcript Collection Process

The transcript collection process focused on converting spoken content from educational videos into written text, enabling detailed analysis of the video's content quality and coherence. This process facilitated the extraction of valuable text data for further evaluation and research.

Figure 3.3: Transcript Extraction Process through Whisper API

3.6.2 Downloading of Educational Videos

YouTube was selected as the primary digital educational platform for downloading of educational videos. All videos were downloaded in **.mp4 format** with a resolution of **720 pixels**, ensuring only high-quality video files are selected for further processing. Videos were downloaded using Internet download manager (IDM) which facilitates efficient downloading in high quality formats.

3.6.3 Extraction of Audio from Video Files

Python library **moviepy** was utilized to extract audio from the video files. The extracted audio was saved in **.mp3** format in the designated folder.

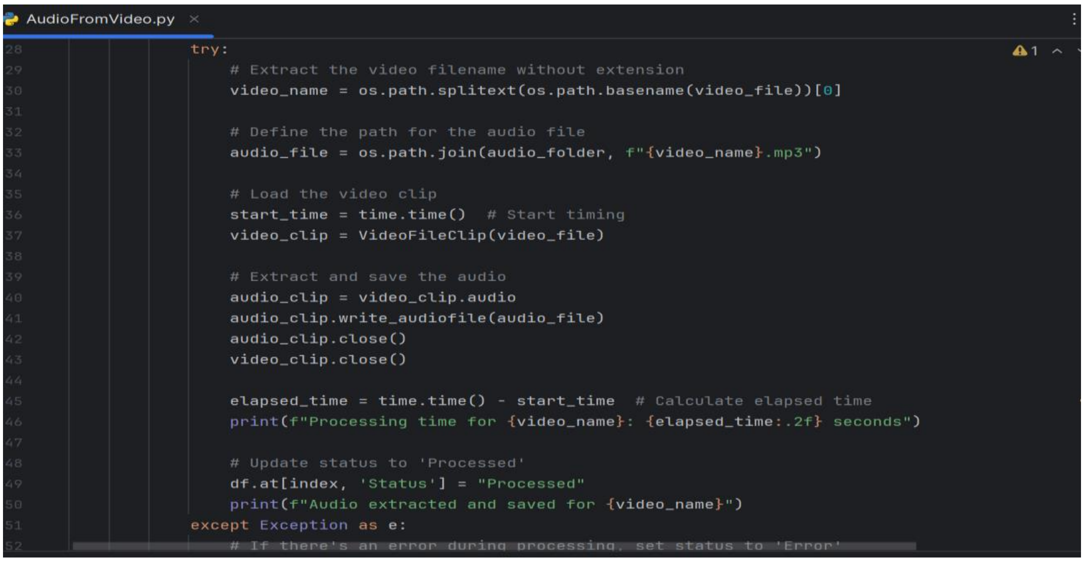


Figure 3.4: Code Snippet for Conversion of Audio from Video

3.6.4 Extraction of Transcript from Audio Files

Transcript conversion process was accomplished using **OpenAI's Whisper AP I**, a cutting-edge tool for speech recognition. Whisper supports 97 languages and is capable of transcribing audio content from various sources, making it ideal for educational videos. In our case, we used base model of Whisper. Due to the complexity of the task and the length of the videos, the transcription process took a considerable amount of time. For example, a 30-minute video took nearly an hour to fully transcribe into text. The transcripts were saved as **.txt files** for further analysis.

3.6.5 Feature Extraction from Transcripts

The feature extraction process involved several key steps to transform raw text into meaningful metrics that reflect the quality and characteristics of the content. Before extracting features, preprocessing on the text was performed to remove noise and focus on the most relevant details of the text. The main steps included were:

- Tokenization
- Stopword Removal
- Lemmatization

After completing the preprocessing steps, the following features were extracted for the creation of the dataset.

3.6.6 Coherence Metrics

In the context of this research, Coherence refers to the logical flow of information throughout an educational video. Coherence of the content was evaluated at two levels – local coherence and global coherence. This process was crucial to assess overall quality of the transcript which forms a key role of multimodal analysis of this study.

- **Local Coherence** Refers to smooth transition between small sections of the transcript, typically in consecutive sentences or short segments. It was calculated using a sliding window technique where **cosine similarity** between sentence embeddings was employed to measure the semantic closeness between adjacent sentences. The local coherence formula used in this analysis:

$$\text{Local Coherence} = \frac{\sum \text{Cosine Similarity between sentence pairs}}{\text{Total number of comparisons}}$$

- **Global Coherence** It measures the overall consistency of the entire transcript. It was really important to ensure that the educational video maintained a clear and focused message throughout its duration. Global coherence was assessed by comparing the **sentence embeddings** of all possible sentence pairs. The global coherence formula used in this analysis:

$$\text{Global Coherence} = \frac{\sum \text{Cosine Similarity between all sentence pairs}}{\text{Total number of sentence pairs}}$$

- **Overall Coherence** to provide a holistic measure of coherence, combined local and global coherence scores into a single metric. The overall coherence captured both short term logical flow (local coherence) and long-term narrative consistency (global coherence), which provided balanced view of the transcript's structure. Coherence ensured that the content is logically structured, which directly impacts learners' engagement and comprehension. The formula for overall coherence used in this analysis:

$$\text{Overall Coherence} = \frac{\text{Local Coherence} + \text{Global Coherence}}{2}$$

3.6.7 Engagement Metrics

These metrics measured how well the content promoted active participation, critical thinking, and sustained learner interest throughout the video. High engagement ensures learners actively participate in the learning process, rather than passively receiving information.

- **Total Sentences** The total number of sentences were calculated which helped in understanding the overall length and structure of the transcript and it also served as the base denominator for other engagement metrics.
- **Question Count** When it comes to educational content, questions play a pivotal role in making the content interactive. In this study, identifying questions was essential to gauge how much the transcript was designed to engage learners in a

dialogue. This feature was calculated by checking the number of sentences ending with a question mark in the transcript.

- **Prompt Count** A prompt is a statement designed to stimulate thought, encourage active engagement, or provoke a response from learners. Prompt Count measured the number of sentences that included action-oriented phrases like "follow along" or "can you imagine" designed to encourage learners to participate actively or think critically. A comprehensive list of prompt phrases was collected from various sites: -
  - https://wisernotify.com/blog/call-to-action-phrases/
  - https://www.espressoenglish.net/question-words-100-example-questions/
  - https://helendoron.com/english/30-phrases-for-encouraging-someone-in-english/
  - https://parade.com/1211362/marynliles/hypothetical-questions/

Table 3.2 List of Prompt Phrases

please	join	reflect on	explore the idea of	could you
you can	check out	how would it be if	what would it be like if	envision
try	don't forget	how about trying	imagine the impact of	what are some ways to
discover	remember to	i suggest	think about how	in what ways
explore	sign up for	i recommend	reflect on how	how could
join	follow along	why don't you	consider the possibility	what are the benefits of
check out	give it a shot	it might be a good idea to	how could we	how might
don't forget	take a look at	suppose	what would happen if	what changes could be made
remember to	watch	let's assume	have you considered	how would this affect
sign up for	listen to	imagine if	what if we	consider the effects of
follow along	what do you think	consider a scenario where	how can we improve	what are some potential
give it a shot	how about	what might be the outcome	what if	how might this influence
take a look at	why not	could you	how might	what are the implications
please	join	reflect on	explore the idea of	could you
you can	check out	how would it be if	what would it be like if	envision

try	don't forget	how about trying	imagine the impact of	what are some ways to
discover	remember to	i suggest	think about how	in what ways
explore	sign up for	i recommend	reflect on how	how could
join	follow along	why don't you	consider the possibility	what are the benefits of
check out	give it a shot	it might be a good idea to	how could we	how might
don't forget	take a look at	suppose	what would happen if	what changes could be made
remember to	watch	let's assume	have you considered	how would this affect
sign up for	listen to	imagine if	what if we	consider the effects of
follow along	what do you think	consider a scenario where	how can we improve	what are some potential
give it a shot	how about	what might be the outcome	what if	how might this influence
take a look at	why not	could you	how might	what are the implications
please	join	reflect on	explore the idea of	could you
you can	check out	how would it be if	what would it be like if	envision
try	don't forget	how about trying	imagine the impact of	what are some ways to
discover	remember to	i suggest	think about how	in what ways

- **Question and Prompt Proportions** Two Features, Question proportion and prompt proportion were calculated by dividing the number of questions and prompts by the total sentences in the transcript. These metrics indicated how often the instructor engaged with learners during the lecture.

$$\text{Question Proportion} = \frac{\text{Number of Questions}}{\text{Total Sentences}} \qquad \text{Prompt Proportion} = \frac{\text{Number of Prompts}}{\text{Total Sentences}}$$

- **Engagement Score** This feature was calculated to reflect overall interactivity of the transcript. Engagement score provided a useful indicator of the content's potential to keep learners mentally active and encouraged engagement through questions and prompts.

$$\text{Engagement Score} = \text{Question Proportion} + \text{Prompt Proportion}$$

3.6.8 Relevance Metric

Relevance is evaluated to ensure that the transcript remained aligned with the intended educational topic. The relevance score was calculated by comparing the TF-IDF representations of the transcript and reference text (educational video title) using cosine similarity.

3.6.9 Readability Metrics

Readability is the measure of how easier a piece of text is. Educational videos must balance complex content with accessibility. These metrics help educators to evaluate how easy or difficult educational video transcript is for learners to understand. In this research following key readability metrics are calculated: -

- **Flesch Reading Ease** Transcript readability is evaluated using this metric on a 100-point scale, with higher scores indicating easier comprehension. It assesses the number of syllables per word as well as sentence length. The Flesch Reading Ease can be computed using the formula: -

$$206.835 - 1.015\left(\frac{\text{Total Words}}{\text{Total Sentences}}\right) - 84.6\left(\frac{\text{Total Syllables}}{\text{Total Words}}\right)$$

Table 3.3 Flesh Reading Ease Score Range

Score Range	Interpretation	Score Range	Interpretation
90-100	Very easy to read. Easily understood by 11-year-olds.	50-59	Fairly difficult. Suitable for high school graduates.
80-89	Easy to read. Conversational English, suitable for 12-year-olds.	30-49	Difficult. Best understood by college students.
70-79	Fairly easy to read. Understood by 13-15-year-olds.	0-29	Very difficult. Best understood by university graduates or professionals.
60-69	Standard. Plain English, understood by 16-17-year-olds.		

- **Gunning Fog Index** This metric counts the quantity of polysyllabic words (words with three or more syllables) and the ratio of words to sentences to determine how tough a transcript is. The Gunning Fog Index is computed using the following formula:

$$0.4\left(\left(\frac{\text{Total Words}}{\text{Total Sentences}}\right)+100\times\frac{\text{Complex Words}}{\text{Total Words}}\right)$$

Table 3.4 Gunning Fog Index Score Range

Score Range	Interpretation	Score Range	Interpretation
6-7	Very easy to read. Easily understood by 11-12-year-olds.	13-15	Difficult to read. Best understood by college students.
8-9	Easy to read. Suitable for middle school students.	16-17	Very difficult. Best understood by university graduates.
10-12	Fairly difficult. Suitable for high school students.	18+	Extremely difficult. Best understood by professionals.

- **SMOG Index (Simple Measure of Gobbledygook)** Estimates the years of education needed to understand a piece of text mostly focusing on the number of complex words. Higher score indicates more complex text. The formula to calculate SMOG Index is: -

$$1.0430\times\sqrt{\text{Number of Polysyllabic Words}\times\frac{30}{\text{Total Sentences}}}+3.1291$$

Table 3.5 SMOG Index Score Range

Score Range	Interpretation	Score Range	Interpretation
1-5	Very easy to read. Suitable for early grade levels (e.g., 4th grade).	11-15	Difficult to read. Suitable for high school students (grades 11-12).
6-10	Fairly easy to read. Suitable for middle school students (grades 6-8).	16-20	Very difficult. Suitable for college students and professionals.
21+	Extremely difficult. Best understood by advanced readers.		

- **Automated Readability Index (ARI)** Based on the average number of characters per word and sentence length, this measure assesses the readability of transcripts and provides a grade level that indicates the level of comprehension required for understanding. The Automated Readability Index (ARI) is computed using the following formula:

$$4.71\left(\frac{\text{Characters}}{\text{Total Words}}\right)+0.5\left(\frac{\text{Total Words}}{\text{Total Sentences}}\right)-21.43$$

Table 3.6 ARI Index Score Range

Score Range	Interpretation	Score Range	Interpretation
5	Very easy to read. Suitable for early grade levels (e.g., 1st-5th grade).	11-15	Difficult to read. Suitable for high school students (grades 11-12).
6-10	Easy to read. Suitable for middle school students (grades 6-8).	16-20	Very difficult. Suitable for college students and professionals.
21+	Extremely difficult. Best understood by advanced readers.		

- **Dale-Chall Readability Score** This statistic helps assess the accessibility of content by offering insightful information about how effectively educational transcripts are understood by various educational levels. The □ Dale-Chall Readability is computed using the following formula:

$$\text{Dale-Chall Readability Score}=0.1579\times\left(\frac{\text{Difficult Words}}{\text{Total Words}}\right)+0.0496\times\left(\frac{\text{Total Words}}{\text{Total Sentences}}\right)+3.6365$$

Table 3.7 Dale-Chall Readability Score Range

Score Range	Interpretation
4.0 - 5.0	Easily understood by 4th-grade students, indicating very accessible text.
5.1 - 6.0	Suitable for 5th-6th grade students, with slightly increased complexity.
6.1 - 7.0	Appropriate for 7th-8th grade students, indicating moderate difficulty.
7.1 - 8.0	Suitable for high school students, with higher complexity.
8.1 - 9.0	Appropriate for college students, with complex language and structure.
9.1 - 10.0	Understandable by readers with some college education, indicating advanced complexity.
10.1 and above	Suitable for readers with higher education levels, such as professionals.

- **Linsear Write** This metric estimates the grade level required to comprehend the material in order to assess the readability of school transcripts. It evaluates the proportion of easy to difficult words in a transcript and is helpful in figuring out how complex a text is meant for learners.

$$\text{Linsear Write Formula} = \left(\frac{\text{Number of Easy Words}}{\text{Total Words}} + 2 \times \frac{\text{Number of Hard Words}}{\text{Total Words}}\right)$$

Table 3.8 Linsear Write Score Range

Score Range	Interpretation
1 - 5	Very easy to comprehend, suitable for elementary learners.
6 - 7	Fairly easy, understandable for middle school students.
8 - 9	Moderately complex, appropriate for high school students.
10 - 12	Complex, requiring college-level education.
13+	Very complex, suitable for professionals or advanced learners.

- **Coleman-Liau Index** This readability metric is used to determine the grade level needed to comprehend a certain book. It is different from previous readability formulas in that it emphasizes letters rather than syllables, making it a useful tool for evaluating digital writing.

  The formula for computing Coleman-Liau Index is: -

$$\text{Index} = 0.0588 \times \left(\frac{\text{Letters}}{\text{Words}} \times 100\right) - 0.296 \times \left(\frac{\text{Sentences}}{\text{Words}} \times 100\right) - 15.8$$

3.6.10 Vocabulary Richness

In order to assess the diversity of words used in the transcript, vocabulary richness was calculated using the following formula.

Ratio of Type-Token (TTR) = Unique Words / Total Words

3.6.11 Finalized Features of Transcript Dataset

Table 3.9 Features of Transcript Dataset

Metric Type	Features
Engagement Metrics	✓ Total Sentences ✓ Question Count ✓ Prompt Count ✓ Question Proportion ✓ Prompt Proportion ✓ Engagement Score
Coherence Metrics	✓ Local Coherence

	✓ Global Coherence ✓ Overall Coherence
Similarity Metrics	Similarity Score
Readability Metrics	✓ Flesch Reading Ease ✓ Gunning Fog ✓ SMOG Index ✓ Automated Readability Index ✓ Dale-Chall Readability Score ✓ Linsear Write Formula ✓ Coleman-Liau Index ✓ Readability Consensus
Vocabulary Metrics	✓ Total Words ✓ Unique Words ✓ Vocabulary Richness

3.7 Scoring and Annotation of Educational Videos

A very thoughtful approach for annotating target variable of both metadata and transcript datasets was made which would ultimately act as a benchmark for evaluation of predictions from machine learning and deep learning models. Survey based method and statistical techniques were used for the scoring of educational videos.

3.7.1 Survey based Annotation for User Feedback

A google survey form was developed and distributed to students, teachers and to general public. There were 1050 educational videos in the dataset from which 100 x videos were selected randomly for the survey. Each participant was asked to watch minimum 20 videos and that too in multiple sessions and provide a rating on a scale of 1 to 5. This provided a subjective evaluation of video quality based on personal perception and engagement. This survey will be used to compare the results of predicted models scores with human feedback. By comparing the human ratings with the model predictions, it might be possible to validate the performance and refine the models for achieving improved results. This is the link of the survey form.

https://docs.google.com/forms/d/e/1FAIpQLSfx3lSkHbYpwK5UT34ys7f8rzSCldJHAJJz4Ru3D5bp319nEw/viewform?usp=sf_link

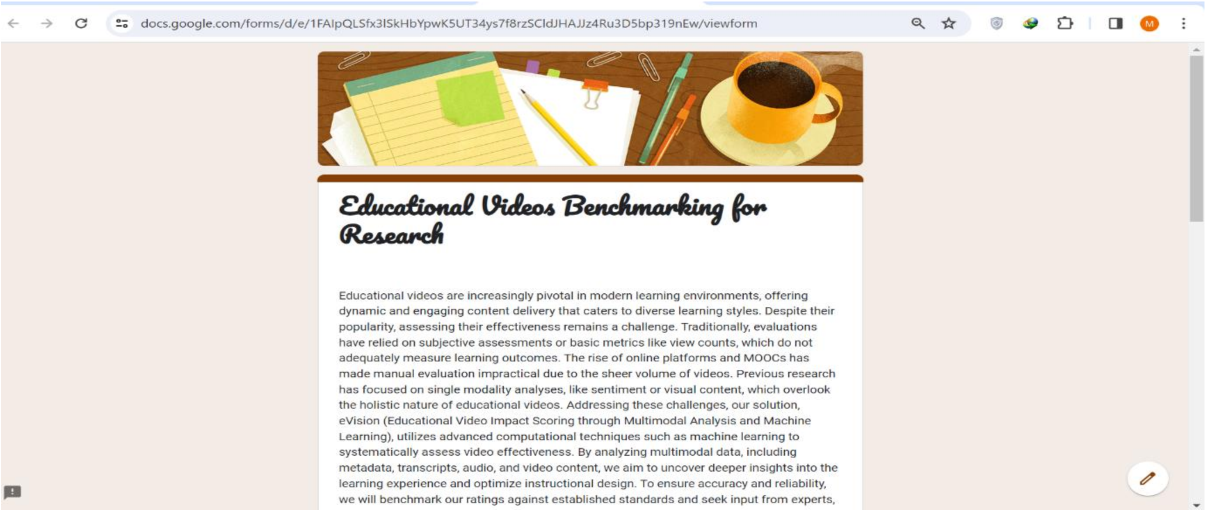


Figure 3.5: Front Page of Google Survey Form

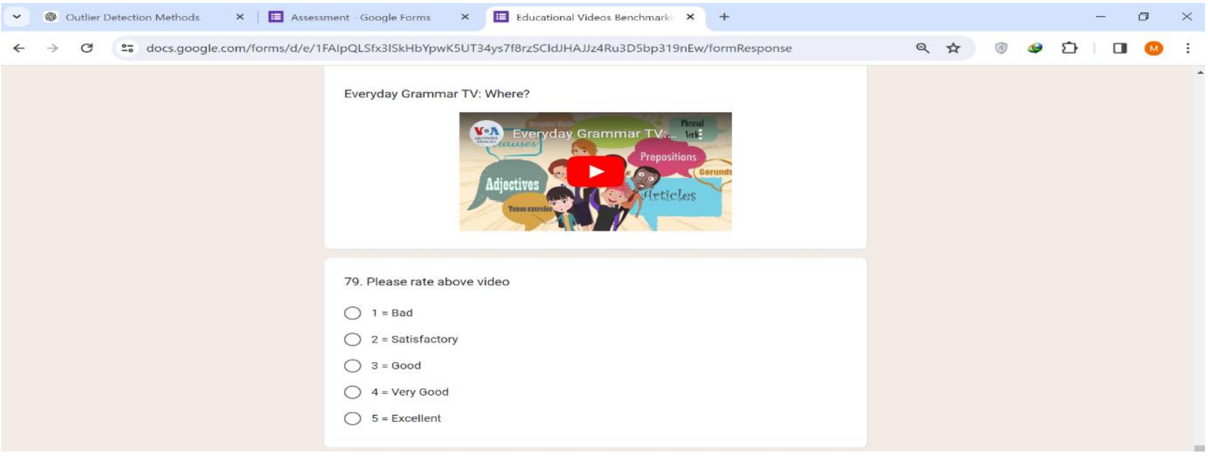


Figure 3.6: Interface for Viewing and Rating Educational Videos

3.7.2 Statistical Based Annotation

After collecting subjective ratings through Google survey for the 100 videos, it was very much clear that manually watching all the remaining 950 videos and then rate them was not possible. So, for that a comprehensive scoring mechanism is designed by using statistical techniques without relying on the subjective human input. Numerical features of both datasets such as view count, comment count, relevance metric formed the basis for a structured scoring system using statistical technique called percentiles.

- **Percentile based Method** This statistical technique is used to rank and categorize data points based on their position within the dataset. It tells what percentage of data lies below a given value. For example, if we are calculating the 80^{th} percentile of a feature, it means that 80% of data points are below this calculated value. To calculate the percentile for a given data point following formula is used: -

$$P = \left(\frac{n}{N}\right) \times 100$$

- **P**: Percentile rank of the data point.
- **n**: The number of data points below the value you're calculating the percentile for.
- **N**: The total number of data points in the dataset.

- **Metadata Scoring Mechanism** Numerical features such as view count, comment count and engagement rate are used for providing a score to an educational video using percentile method. This technique categorized features into different categories. For example, Viewcount feature of metadata dataset was divided into five ranges using percentile based method which are: -
  - Videos in the **80^{th} Percentile and above** are given a score of **5** (indicating very high viewership).
  - Videos in the **60th to 80th percentile** are given score of **4**.
  - Videos in the **40th to 60th percentile** received a score of **3**.
  - Videos in the **20th to 40th percentile** received a score of **2**.
  - Videos **below the 20th percentile** received a score of **1** (indicating low viewership).

Similarly, this technique was applied to all the numerical features of this dataset except the duration feature. The average attention span of a user is limited and shorter videos are easier to watch in totality. So keeping this in mind, the scoring criteria for **duration** feature is defined in following categories: -

- Videos **less than 6 minutes** received the highest score **5 (Highly Engaging)**
- Videos **between 6 to 10 minutes** received a score of **4**
- Videos **between 10 to 15 minutes** received a score of **3**
- Videos **between 15 to 20 minutes** received a score of **2**

- Videos **longer than 20 minutes** received the lowest score **1 (Low Engaging)**

After calculating the scores of all the features in the range of 1 to 5, target variable for an educational video is calculated using two ways. First method is, by averaging the scores of all the features. In addition, weighted average score is also calculated to give more importance to certain features. For example, view count and like count carry more weight (0.15) than comment to view ratio (0.02). This ensured that key features with high relevance influenced the overall score more heavily. The target variable calculated is a continuous value ranging from 1 to 5.

- **Transcript Scoring Mechanism** Thorough scoring mechanism is also curated for the transcript dataset. A combination of percentile-based method and specific range-based method for the readability assessment metrics is applied to calculate the overall score of an educational video transcript. For example, **Prompt count** feature score is calculated using the same percentile-based method used in metadata. For calculating score of the readability assessment metrics, scale range of that specific metric is used. Let's give an example of Gunning Fog index how its score is calculated: -
  - Fog Index **0–6** received a score of **5 (Very Easy to Read)**
  - Fog Index **7–10** received a score of **4 (Easy to Read)**
  - Fog Index **11–14** received a score of **3 (Moderate Reading Difficulty)**
  - Fog Index **15–18** received a score of **2 (Difficult to Read)**
  - Fog Index **19+** received a score of **1 (Very Difficult to Read)**

  Unlike, Metadata dataset, each feature contributes unique insights into the transcript's quality and engagement, providing a holistic view of the content. Therefore, final rating is only derived by averaging all the feature scores.

Chapter 4 - Implementation of Proposed Methodology

4.1 Overview

In this chapter, implementation of proposed model of eVISION is described in detail which includes environment used for this research work, data preprocessing techniques applied to both metadata and transcript datasets. All the machine learning and deep learning regression models utilized for this research are described in details followed by evaluation of parameters.

4.2 Environment

All the experiments are executed using language Python 3.9.13 using PyCharm IDE 2023.3.4 (Professional Edition). 8 GB RAM with windows 10 on 12th Gen Intel(R) Core(TM) i5-1235U 1.30 GHz is used in this research work.

4.3 APIs Utilized for Research

In this research, several APIs are used to handle different tasks essential for the analysis of educational videos. These APIs facilitated the extraction of data, conversion of media, and transcription of audio, each contributing to the overall workflow.

- **Google YouTube API**: Retrieves videos statistics from YouTube.
- **Moviepy**: Converts videos to audio format.
- **OpenAI's Whisper API**: Speech to text model that transcribes audio to text.

4.4 Data Preprocessing

To get the data ready for analysis and model training, a number of preprocessing steps were implemented after the generation of the transcript and metadata datasets. Some features in both datasets had missing values which were dealt accordingly. Outliers were found and dealt with to preserve data integrity and avoid analytical distortion, either by deleting or capping them. In order to normalize the data of both the datasets min max normalization technique is applied to scale the data in a specific range.

4.5 Min-Max Normalization

Min-Max scaling is a widely used scaling technique in ML. This normalization process guarantees that all features are uniformly scaled from 0 to 1 and prevents certain features from exerting excessive influence on the learning procedure. By standardizing the features, convergence of optimization algorithms is improved, facilitating faster and more stable training.

4.6 Predictive Modeling

Due to continuous nature of target variable which represents the video scores, regression algorithms are used as they can predict continuous range of values. The goal is to develop a model capable of predicting video quality based on the metadata and transcript features. Below is a detailed description of the models used.

4.6.1 Linear Regression

Linear regression (LR) is a basic and primary algorithm in machine learning its simple model which provides an approximation of the relationship between dependent variable values (target attributes) and independent variables. Fits a straight line through the data that models the relationship between one dependent variable (target) and one or more independent variables (features). In essence, the goal is to identify a line where the difference (or error) between predicted and actual values of target variable is minimum.

4.6.2 Decision Tree

Decision Tree (DT) Regression is a non-linear algorithm that is used for predictive modeling and serves as a splitting mechanism for the data into subsets based on the most important features. It functions by making a tree-like model of decisions and their possible outcomes. For regression, a continuous output is predicted as the system learns from the data decision rules. The tree consists of nodes, where each node represents a decision based on a feature value. The leaf nodes represent the predicted outcome. It is suitable for handling complex, non-linear relationships between features and continuous target variables.

4.6.3 Random Forest Regressor

The Random Forest Regressor (RFR) is a well-known machine learning algorithm that is widely used for regression problems, where the predicted variable is an arbitrary target variable which represents a continuous number. It works based on an ensemble learning method, forming a lot of decision trees in the training phase and combining their outputs to improve the accuracy and stability of the predictions. Each tree is made using a randomly selected subset of the training data and features. The predictions of the entire forest (consisting of each tree) are then averaged out to give us the final prediction. This helps to mitigate overfitting: by randomizing both the data and the feature subsets each tree is built on, we ensure that no single tree comes to dominate the decision-making process so that it works only on the specific feature configuration and training sample it was built on. The model generally fares well against noise, and is amenable to big data with high-dimensional features.

4.6.4 Ridge Regression

Ridge Regression (RR) is a variant of linear regression which deals with the problem of multicollinearity (when the independent variables are highly correlated) by adding a regularization term (alternatively called penalty term) to the cost function. This helps to prevent the model coefficients from growing unnecessarily large, thus reducing overfitting instead of minimizing the cost function. The strength of the penalty applied is controlled by the parameter alpha or lambda. The aim of Ridge Regression, as its name suggests, is to optimize the penalized total sum of squares errors: The second part of the equation, which multiplies the squared coefficient by a tuning parameter λ, is the penalty function. This penalty aids generalization in two ways: Firstly, by limiting absurdly large coefficients. Ridge is particularly useful when there are many features, or in the presence of noise, to reduce overfitting, as the model is less complex.

4.6.5 SVR

Support Vector Regression (SVR) also known as SVM Regression, is the supervised machine learning technique which estimates any continuous values with the aim of finding the best hyperplane in the feature space of higher dimensions. To deal with complex problems, SVR employs the "kernel trick," which enables the transformation of the input feature space to handle nonlinear relationships. SVR proceeds by optimizing the boundary with regard to a given tolerance. SRV is the subclass of regression analysis that deals with data regarding forecasting trending investments including the pricing of stocks through prediction. It is a highly flexible and powerful algorithm capable of working in regression mode and is applicable on nonlinear as well as high dimensional data.

4.6.6 XGBoost Regression

XGBoost (XGB), short for Extreme Gradient Boosting, is a sophisticated and effective machine learning model, mainly applied in regression and classification problems. It belongs to the family of boosting algorithms, which gradually adds up ‘weak’ predictors, usually decision trees, in the hopes of forming a singular, more competent than any of the individual components, informing mechanism. XGBoost, on the other hand, leverages regularization in an attempt to address the issues of the gradient boosting approach. XGBoost once again optimizes the model, however, and uses a different approach whereby in every step an error is made by the model, there is a corresponding determination of the loss function. XGBoost is very efficient and can easily scale up to large data sets that allow its deployment in real life applications such as time series forecasting, recommendation systems and ranking tasks.

4.6.7 Feedforward Neural Network (FNN)

A Feedforward Neural Network (FNN) is one of the most basic artificial neural networks and is employed for regression or classification purposes. Feed Forward Layer Neural Network is referred to as FNN because the information in an FNN moves only in one direction which is from the input layer to the hidden layers till it reaches the output layer. This model is particularly effective in learning complex patterns through a series of weighted connections between neurons. By adjusting these weights during training, the network learns to minimize the error between predicted and actual outputs, using techniques like backpropagation and gradient descent.

4.7 Evaluation Parameters

In this work, the performance of regression models was assessed using three crucial assessment metrics: Mean Squared Error (MSE), Mean Absolute Error (MAE), and R-Squared (R^2). These metrics evaluate the degree of agreement between the predicted and actual values and provide insight into the accuracy and reliability of the regression models.

4.7.1 Mean Squared Error (MSE)

It measures the average squared difference between the predicted values and the actual values, penalizing larger errors more heavily. Goal is to minimize MSE near to zero to ensure predictions are close to actual values.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and n is the number of observations.

4.7.2 Mean Absolute Error (MAE)

The average absolute difference (MAE) between the actual and anticipated values is calculated. It provides a sense of the average deviation between the predictions and the actual numbers. Similarly, like MSE the aim is to minimize MAE near to zero to ensure accurate predictions.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

4.7.3 R-Squared (R^2)

Indicates how well the model explains the variance in the data. An R^2 value closer to 1 indicates that the model explains most of the variance in the target variable.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where $\bar{y}$ is the mean of the actual values.

4.8 Summary

Implementation of complete proposed model is discussed in this chapter. Tools, APIs and data preprocessing techniques are discussed and how these techniques enhanced the results are discussed in details. Whereas all the six ML regression models are explained followed by evaluation parameters are highlighted based on which all the models were examined efficiently.

Chapter 5 - Experimental Results and Discussion

5.1 Overview

The results section of this thesis presents the findings from the analysis of educational videos, utilizing both metadata and transcript-based features to predict video quality scores. The aim of this research was to develop a model which accurately evaluates educational content based on various features. A range of regression-based methods including LR, RGR, DT, RFR, SVR, XGB and FNN were employed to check quality of the educational videos. The methodology and models used in this research provided promising results in evaluating the quality of educational videos, helping to assess and rank them effectively. The performance of these models was assessed using metrics such as MAE, MSE and R-Squared (R^2) to gauge their accuracy and reliability.

5.2 Performance of Prediction Models on Metadata Dataset

As target variable of Metadata dataset was annotated using average and weighted average of all the features scores. So, therefore experiments are performed on both these target variables to evaluate which statistical technique is better after applying different prediction models to predict the values.

5.2.1 Model Evaluation Using Feature Average Target Variable

Firstly, experiments are performed on the metadata dataset which has annotated target variable calculated by averaging all the feature scores. Table 5.1 indicates the results of six machine learning prediction models and using deep learning feed forward neural network.

Table 5.1 Performance of Prediction Models on Metadata Average Based Method

Prediction Model	MSE	MAE	R^2
LR	0.44	0.48	0.08
DT	0.13	0.03	0.90
RF	0.09	0.02	0.97

SVR	0.58	0.51	0.02
RR	0.46	0.33	0.38
XGB	0.08	0.06	0.98
FNN	0.05	0.14	0.97

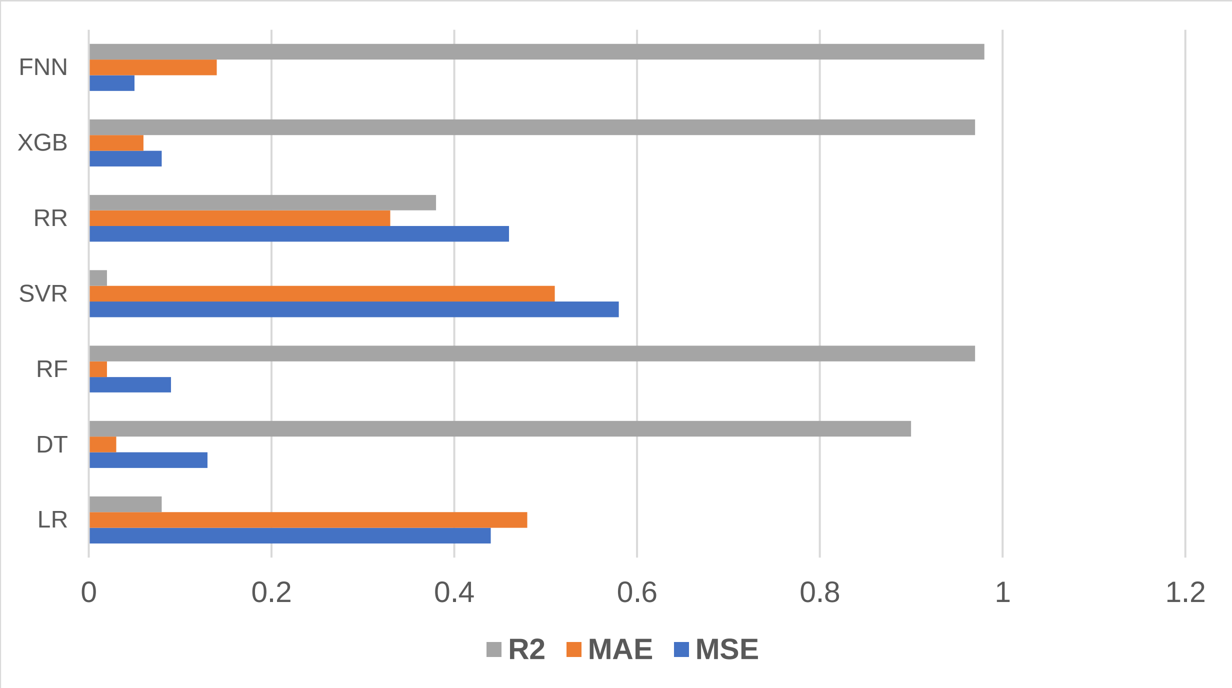


Fig 5.1 Graphical View of Prediction Model - Average Target Variable

XGB performed the best overall with the lowest MSE of 0.08, MAE 0.06 and highest R^2 as 0.98. FNN also performed well with an MSE of 0.05 and R^2 as 0.97.

5.2.2 Model Evaluation Using Weighted Feature Scores

Experiments are then performed on the metadata dataset which has annotated target variable calculated by taking weighted mean of all the feature scores. Table 5.2 indicates the results of six machine learning prediction models and using deep learning feed forward neural network.

Table 5.2 Performance of Prediction Models on Metadata – Weighted Average Score

Prediction Model	**MSE**	**MAE**	**R^2**
LR	0.57	0.80	0.05
DT	0.12	0.03	0.94

RF	0.09	0.06	0.97
SVR	0.71	0.74	0.07
RR	0.58	0.53	0.31
XGB	0.08	0.07	0.98
FNN	0.07	0.05	0.99

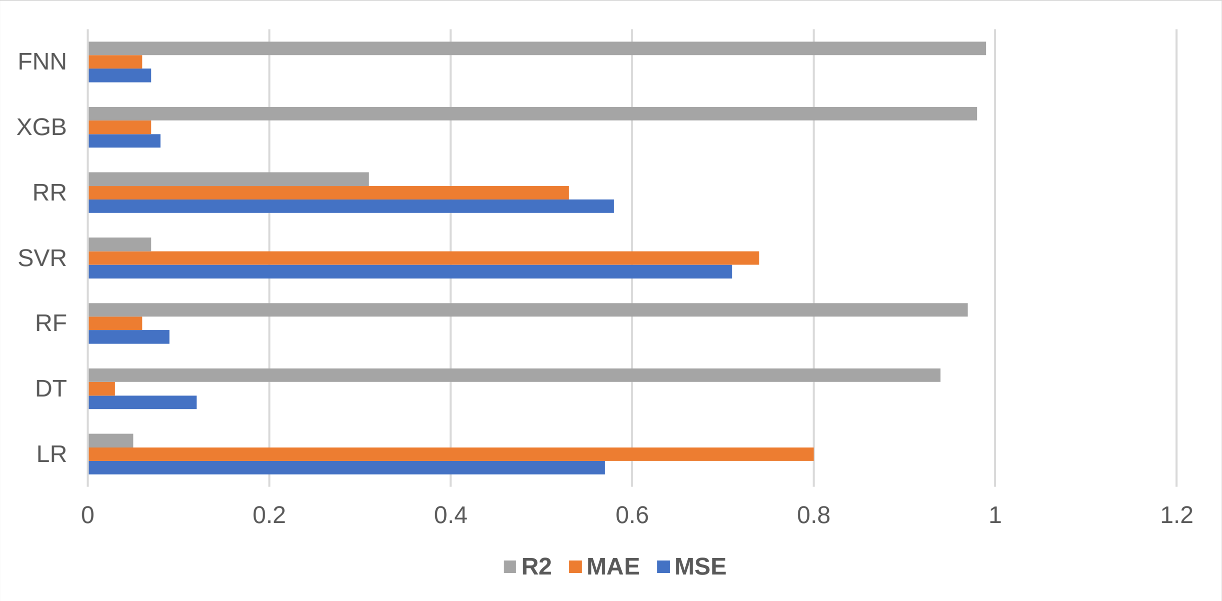


Fig 5.2 Graphical View of Prediction Model - Average Target Variable

With the lowest MSE (0.07), MAE (0.05), and greatest R^2 (0.99) of all the prediction models applied to the weighted average target variable, FNN performed the best. With an MSE of 0.08 and R2 of 0.98, XGBoost likewise demonstrated outstanding performance; but, in terms of accuracy, the FNN slightly surpassed it.

5.2.3 Performance Analysis: Average vs. Weighted Average

The weighted average-based approach using FNN performs the best overall, providing the highest R^2, a sign of higher prediction accuracy, and the lowest MSE and MAE.

Table 5.3 Comparison of Best Performing Models on Metadata

Model	Method	MSE	MAE	R^2
XGB	Average-Based	0.08	0.01	0.98
FNN	Average-Based	0.05	0.14	0.97

XGB	Weighted Average	0.08	0.06	0.98
FNN	Weighted Average	0.07	0.05	0.99

5.3 Performance of Prediction Models on Transcript Dataset

The experimental results for the transcript dataset are summarized below after applying various prediction-based models. Additionally, FNN is also used for prediction.

Table 5.4 Performance of Prediction Models on Transcript Dataset

Model	MAE	MSE	R^2
LR	0.17	0.04	0.71
DT	0.15	0.04	0.71
RF	0.11	0.02	0.87
SVR	0.28	0.11	0.24
Ridge	0.17	0.04	0.71
XGB	0.11	0.02	0.84
FNN	0.10	0.02	0.88

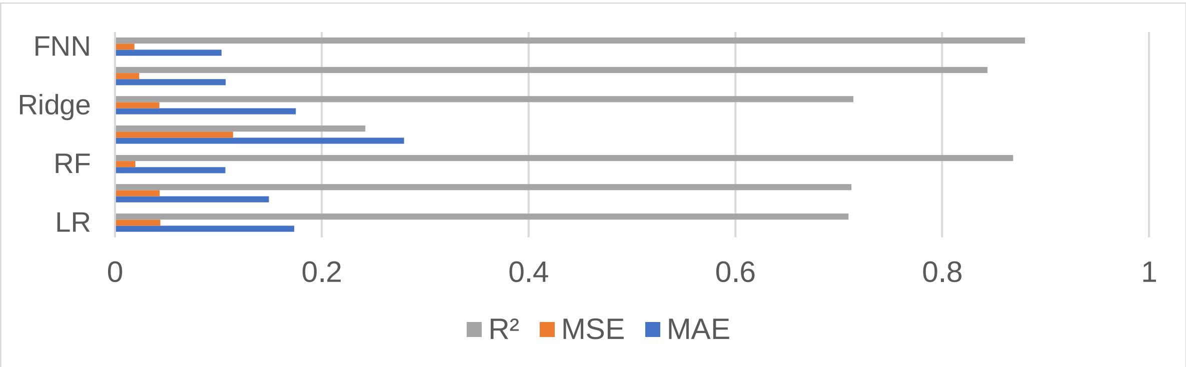


Fig 5.3 Graphical View of Transcript Dataset Accuracy

The Feedforward Neural Network is the best-performing model in this comparison. It achieves the lowest MAE and MSE, indicating the smallest prediction errors, and its R^2 value of 0.88 shows that it explains about 88% of the variance in the target variable.

5.4 Performance of Best Prediction Model on Human Feedback

We compared our best prediction models to predict values of 100 educational videos and then compared the results with the human feedback which was taken through Google survey form. Table 5.5 shows comparison of human feedback with our best prediction models. Here, we used prediction model of metadata to compare the results with the feedback of users.

Table 5.5 Comparison of Prediction Models with Human Feedback

Metric	Human Feedback vs. FNN	Human Feedback vs. XGB
MSE	0.2632	0.2775
MAE	0.3152	0.3287
R^2	0.6142	0.6029

Subjective evaluations impacted by personal prejudices and preferences are a natural part of human feedback. Inconsistencies in feedback may result from this variability, which models must handle in order to produce precise predictions. The fact that FNN's error metrics are lower indicates that it has more successfully adjusted to these arbitrary components, yielding predictions that are more in line with human opinion. Both models, however, are comparatively competent; their performance measures show that they can simulate human feedback rather accurately.

5.5 Summary

In this chapter, performance of various ML models and FNN for predicting scores of educational videos using its metadata and textual analysis is assessed. Best performing models are applied on the benchmark dataset to predict the values and then those values are compared with the human feedback. The results demonstrated reliability of FNN ability to assess educational video quality positioning it as a valuable tool for educational video evaluation.

Chapter 6 - Conclusion and Future Work

6.1 Conclusion and Objective Achieved

The proposed framework has highlighted its potential for evaluating educational video quality, setting a strong foundation for improving video recommendation systems.
This study incorporated range of metadata and transcript features using state of art APIs. Preprocessing techniques were employed to enhance model performance and ensure the reliability of predictions. A hybrid approach was adopted by testing several predictions based models such as XGB, SVR etc. along with FNN for predicting educational video quality score. FNN and XGB models proved to be the best among them who outperformed other models.

6.2 Future Work

In future work, other modalities such as audio and video can be integrated to provide a holistic approach for evaluating educational videos, including analyzing emotions through sound and detecting visual gestures. Additionally, deep learning techniques like LSTM, CNN, and other advanced methods could be employed to further enhance the accuracy and robustness of the results. This multimodal approach will allow for a more comprehensive assessment of educational content, improving the overall evaluation process.

Chapter 7 – References

[1] Sharma, M., & Ankit, P. (2023). Importance of Education in This Challenging World. *SMART MOVES JOURNAL IJELLH*, 11(3), 9-19. https://doi.org/10.24113/ijellh.v11i3.11408

[2] Decuypere, M., Grimaldi, E., & Landri, P. (2020). Introduction: Critical studies of digital education platforms. Methodology of Educational Sciences Research Group, KU Leuven, Leuven, Belgium; Department of Social Sciences, University Federico II, Naples, Italy; Department of Social, Human Sciences and Cultural Heritage, CNR-IRPPS, Fisciano (SA), Italy. Received 14 December 2020; Accepted 15 December 2020.

[3] The Impact of Educational Videos on the Academic Performance of University Students in Distance Learning. (2022). Journal of Positive School Psychology, 6(12), 1233-1249.

[4] UNESCO Institute for Statistics (UIS) (2022). "Out-of-School Children and Youth." Available at: UNESCO

[5] United Nations Educational, Scientific and Cultural Organization (UNESCO) (2021). "Global Education Monitoring Report." Available at: UNESCO GEM Report

[6] Carvalho, H. C. F. B., Dorça, F. A., Pitangui, C. G., Andrade, A. V., Assis, L. P., & Trindade, E. A. C. (2024). Improving the educational experience on YouTube: A machine learning approach to classifying and recommending educational videos. Revista de Gestão e Secretariado – GeSec, 15(4), 01-20.

[7] Andry, J. F., Christianto, K., Lee, F. S., & Loisa, J. (2021). Algorithm of trending videos on YouTube analysis using classification, association and clustering. In Proceedings of the International Conference on Data Science and Engineering (ICoDSE) (pp. 1-6). IEEE. https://doi.org/10.1109/ICoDSE53690.2021.9648486

[8] Amanda, R., & Negara, E. S. (2023). Analysis and implementation machine learning for YouTube data classification by comparing the performance of classification algorithms. Journal of Machine Learning Research, 15(1), 45-60.

[9] Carmichael, M., Reid, A.-K., & Karpicke, J. D. (2023). *Assessing the impact of educational video on student engagement, critical thinking, and learning*. SAGE Publishing. https://us.sagepub.com/sites/default/files/hevideolearning.pdf

[10] Pokharel, R., & Bhatta, D. (2021). Classifying YouTube comments based on sentiment and type of sentence. arXiv preprint arXiv:2111.01908.

[11] De Siva, N.H.T.M., & Rupasingha, R.A.H.M. (2023). Classifying YouTube videos based on their quality: A comparative study of seven machine learning algorithms. Proceedings of the 2023 International Conference on Industrial and Information Systems (ICIIS), 10253580. https://doi.org/10.1109/iciis58898.2023.10253580

[12] Yaradilmis, Y. U., Evren, A. T., & Okkaoglu, M. C. (2023). Evaluation of quality and reliability of YouTube videos on spondylolisthesis. Journal of Physical Medicine and Rehabilitation Sciences, 26(3), 276-281. https://doi.org/10.1016/j.jpmrs.2023.03.001

[13] Kuru, T., & Erken, H. Y. (2020). Evaluation of the quality and reliability of YouTube videos on rotator cuff. Cureus, 12(2), e6852. https://doi.org/10.7759/cureus.6852

[14] Andan, C., & Aydin, M. F. (2022). Evaluation of the reliability and quality of YouTube videos on ovarian cysts. Cureus, 14(3), e22739. https://doi.org/10.7759/cureus.22739

[15] Sentiment analysis for YouTube educational videos using machine and deep learning approaches. Proceedings of the 2022 IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI), 27-29 May 2022. https://doi.org/10.1109/ICETCI55101.2022.9832284

[16] Shoufan, A. (2019). Estimating the cognitive value of YouTube's educational videos: A learning analytics approach. Computers in Human Behavior, 92, 450-458. https://doi.org/10.1016/j.chb.2018.03.001

[17] Mesgar, M., & Strube, M. (2018). A neural local coherence model for text quality assessment. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, 4328-4339. https://doi.org/10.18653/v1/D18-1464

[18] Suresh, V., Gold, A. A., Agasthiya, R., & Chandru, D. (2023). AI-based Automated Essay Grading System using NLP. In *2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 17-19). IEEE. https://doi.org/10.1109/ICICCS56967.2023.10142822

[19] Elizaveta I. Dmitrieva , Ilya V. Grubin. (2021). Assessment of the Text Content Quality of a Distance Technical Translation Course for Graduate Students. In *2021 International Conference on Quality Management, Transport and Information Security, Information Technologies (IT&QM&IS)* (pp. 6-10). IEEE. https://doi.org/10.1109/ICICCS56967.2023.10142822

[20] Shad, M., Yousaf, M. M., & Sarwar, S. M. (2023). Comprehensive Readability Assessment of Scientific Learning Resources. In *IEEE Access*, 11, 53978-53994. https://doi.org/10.1109/ACCESS.2023.3279360

[21] Soeken, M., Abdessaied, N., Allahyari-Abhari, A., Buzo, A., Musat, L., Pelz, G., & Drechsler, R. (2021). Quality Assessment for Requirements based on Natural Language Processing. In *2021 International Conference on Quality Management, Transport and Information Security, Information Technologies (IT&QM&IS) * (pp. 6-10). IEEE. https://www.dfki.de/fileadmin/user_upload/import/7584_2014_fdl_special_session_requirements.pdf