Student Engagement Analysis in Virtual Class Using Natural Language Processing

By Dhanush G

WORD COUNT	2636	TIME SUBMITTED	05-MAR-2024 12:58PM
		PAPER ID	107337789

Student Engagement Analysis in Virtual Class Using Natural Language Processing

Asst. Prof. Balamurugan M [1], Dhanush G [2]

Department of Computer Science and Engineering,
Sona College of Technology, Salem, Tamil Nadu, India

Email: balamurugan.cse@sonatech.ac.in[1], dhanushg.20cse@sonatech.ac.in[2],

Abstract: The shift to digital education has become imperative, particularly during pandemic situations, leading to an increased reliance on online classes. However, teachers face challenges in gauging student involvement and understanding in virtual environments. This research explores the use of NLP techniques, including the BERT model for understanding relationships among words and bag of words for identifying overlapping words between paragraphs, to address these challenges. We propose a method where teachers make notes about lectures and encourage students to take notes in each class. These notes are then analysed using NLP algorithms to trace each student's overall understanding and track how their comprehension evolves over time. The study aims to evaluate how students' understanding and involvement in virtual classes can be enhanced through NLP analysis. We present a portfolio that visualizes the overall understanding graph of each student and their understanding score for each class, providing insights for improving virtual teaching methodologies and student learning outcomes.

Keywords: Digital education, Online classes, Student involvement, Comprehension tracking, Virtual teaching methodologies, Student learning outcomes

I. INTRODUCTION

The way we learn has changed a lot with everything going digital, especially during tough times like the recent pandemic. Online classes have become the new normal, but they come with their own challenges, especially for teachers. One big challenge is figuring out how well students are really understanding and engaging with the lessons in virtual classes. To tackle this challenge, our research dives into the world of Natural Language Processing (NLP), a fancy term for how computers understand human language. We use NLP to analyse the notes students take in online classes, giving us a peek into how well they are understanding and participating in the lessons.

One cool tool we use is called the BERT model. It is like a language detective, sniffing out the meanings behind words in the notes. By using BERT, we can see how much of the teacher's content is getting through to each student. This helps teachers adjust their teaching to make sure everyone is on the same page. Another trick up our sleeve is the bag-of-words technique. This one helps us spot if students are copying from each other by looking at the words they use in their notes. It is like a plagiarism police for online classes, helping keep things fair and honest. To make things even more interesting, we've created a special website for teachers. This website, made with a tool called Stream lit, is like a digital command centre. It stores all the notes and analyses them using an SQLite database. Then, we use cool visual tools like Matplotlib and Plotly to show the data in easy-to-understand graphs and charts.

Our goal is to help teachers better understand how their students are doing in online classes. With our research, we hope to make online learning more engaging and effective for everyone involved, even when we cannot be in the same room together.

II. RELATED WORKS

Suhye Kim et al. investigated elementary students' behaviours in low attention states during online learning using EEG, identifying characteristic movements from webcam videos [1]. Su et al. (2021) introduces a non-intrusive video analytic system utilizing deep learning for real-time monitoring of in-class student learning behaviours, including facial expressions and gestures. Their approach, compatible with edge devices, offers automatic feedback to instructors, enhancing situational awareness and potentially improving learning outcomes in modern educational settings [3]. Li et al. propose an intelligent monitoring algorithm for students' online learning behaviour based on user portraits, utilizing demographic, academic, behavioural, and psychological attributes. This research aligns with our project's objective of understanding student

engagement and comprehension in online classes, potentially informing the development of algorithms for analysing student behaviour and performance using demographic and behavioural data.[10]

Alzahrani et al. (2021) examines various forms of plagiarism, distinguishing between literal and intelligent plagiarism behaviours. Their taxonomy of plagiarism types and survey of detection methods inform our project's development of NLP tools to detect plagiarism in online class notes, complementing our focus on student engagement and comprehension analysis [2].Blanco and Moldovan (2015) propose a semantic logic-based method for determining textual similarity, leveraging logic form transformations and supervised machine learning. Their approach considers semantic structure to improve similarity scores, showing performance enhancements compared to baseline methods and third-party systems. This work aligns with our project's objective of utilizing NLP techniques to analyse textual content in online classes, potentially contributing to the development of more accurate similarity detection algorithms [4].

Zhen et al. propose a novel approach utilizing natural language processing and deep learning to predict academic performance based on dialogue analysis in online classrooms. Their findings highlight the significance of emotional expression and interactive types in distinguishing high- and low-performing students across STEM and non-STEM courses, contributing valuable insights to the field of educational analytics.[5] Aboutaleb et al. propose the BERT BiLSTM-Attention Similarity Model, enhancing the accuracy of question similarity calculation in Question Answering Systems. By leveraging BERT for embedding and BiLSTM-Attention for feature extraction, the model achieves an 84.45% accuracy in determining question similarity, demonstrating its efficacy in improving semantic similarity models for NLP applications.[6]

Tekgöz et al. introduce a novel method for measuring semantic similarity in texts, focusing on sentence and short paragraph comparison. Their approach combines a corpus-based measure of semantic word similarity with a modified Longest Common Subsequence (LCS) algorithm. Evaluation on two datasets demonstrates superior performance compared to existing methods, indicating its potential for applications in textual knowledge representation and discovery in the field of health.[7] Halak and El-Hajjar address the detrimental effects of plagiarism in education and propose two techniques for detection and prevention, emphasizing the importance of fostering creative thinking and maintaining trust between educators and students.[8]Qader et al. provide an overview of the Bag of Words (BoW) method, highlighting its significance, implementation, applications, and challenges. This study could inform our project by offering insights into the use of BoW for text classification, which aligns with our goal of analysing student notes in online classes using Natural Language Processing techniques like BoW to detect plagiarism and assess comprehension.[9]

III. MATERIALS AND METHODS

3.1 Data Collection

We began by collecting student notes from online classes as Real time Dataset. We also use to extract the similar content written by different authors in various platforms using web scraping techniques which is used for training purposes. This involved using the Beautiful Soup module in Python to extract the notes from various online platforms and tools used for classes and note-taking. By gathering notes from diverse sources, we aimed to create a diverse dataset that could provide a comprehensive view of student engagement and comprehension which is shows in below image Figure3.1

Figure: 3.1 Data Collection from Various Sources

3.2 Preprocessing

Before analysing the notes, we pre-processed them using natural language processing (NLP) techniques. This step included removing stop words, tokenization (splitting the text into individual words or tokens), and cleaning the text using NLTK, SpaCy, and regex-based functions. Preprocessing was essential to ensure that the text was in a standardized format and ready for analysis. After these steps we convert the pre-processed data into machine understandable form and so we use

word2vec method in NLP for converting these words into numbers which ML models use to train and validate the dataset. This preprocess steps flow is illustrated in below figure 3.2.

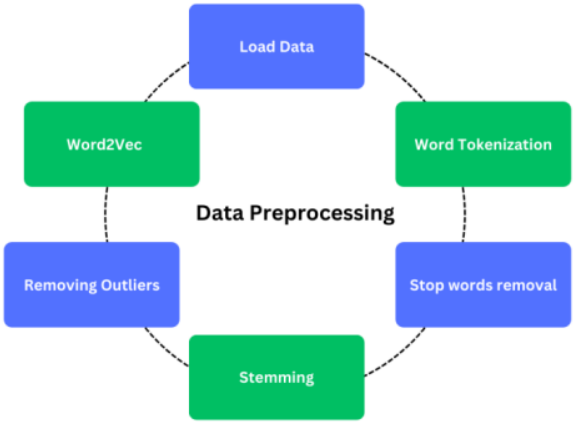

Figure: 3.2 Workflow of Data Preprocessing

3.3 MODEL WORKING

To identify similar paragraphs within the notes, we employed AI models like GPT and BERT. These models can understand the context and meaning of text, allowing us to retrieve paragraphs with similar meanings. This process helped us gain insights into students' comprehension levels and the effectiveness of teaching materials.

3.3.1 BERT MODEL

We use the BERT Model to find similarities between Student Notes with Teacher Notes. BERT Model only processes the embedding vector data. So, in the previous step, we pre-processed the data and converted it into vectors with some pre-processed data. Then we perform the Word embedding to capture the relationship between the words in a sentence. The BERT Model does not perform sequential data processing, instead, it processes the data from both ends to understand the meaning of the sentence. Then we use the any of NLP similarity measures like cosine similarity etc… to calculate the similarity score between two paragraphs means student note and Teacher Notes for understanding how they are like analyse the student engagement in the virtual class which is illustrated in the below figure 3.3.1

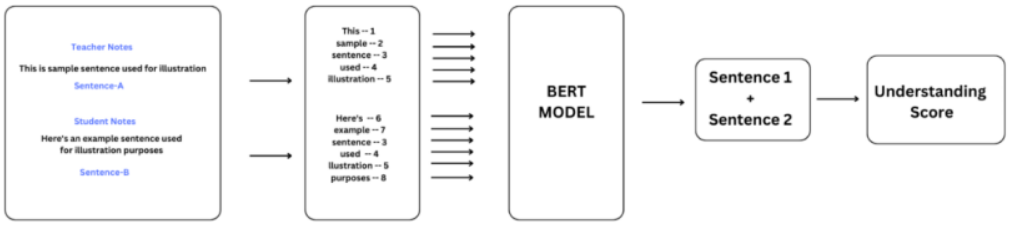

Figure 3.3.1 Working of BERT Model

3.3.2 BOW MODEL

To detect plagiarism among student notes, we employed the Bag-of-Words (BoW) model. This Model also only accepts the vectorized data. So, for this model also we feed only the vectorized data it analyses the by noticing the overlapping words between the notes. We feed student notes to find weather the student copies the content from other student that can be noticed with help of this model which is shown in the below figure 3.3.2

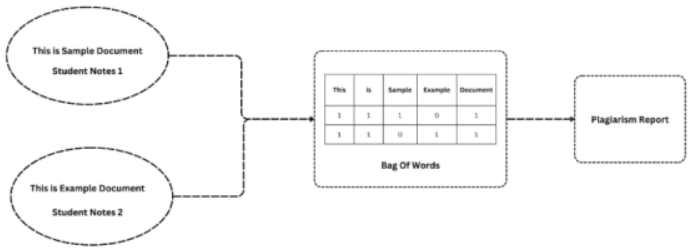

Figure 3.3.2 Working of BoW Model

3.4 VISUALIZATION-PORTFOLIO

This Streamlit application, titled "Student Portal," serves as a multifunctional tool for students. It allows users to log in securely and access various features such as uploading notes, viewing guidelines from teachers, clarifying doubts, and analyzing individual performance through interactive plots showing understanding over time and understanding versus difficulty. Through a user-friendly interface, students can efficiently manage their academic tasks and track their which is shows in figure 3.4.1

The Teacher Portal offers educators a comprehensive platform to manage academic tasks efficiently. With secure login credentials, teachers can upload guidelines, view individual student understanding levels, and analyze overall class performance through interactive visualizations. The portal enhances collaboration by providing fea es for clarifying doubts and offering guidance, fostering a dynamic and supportive learning environment which is shows in figure 3.4.2

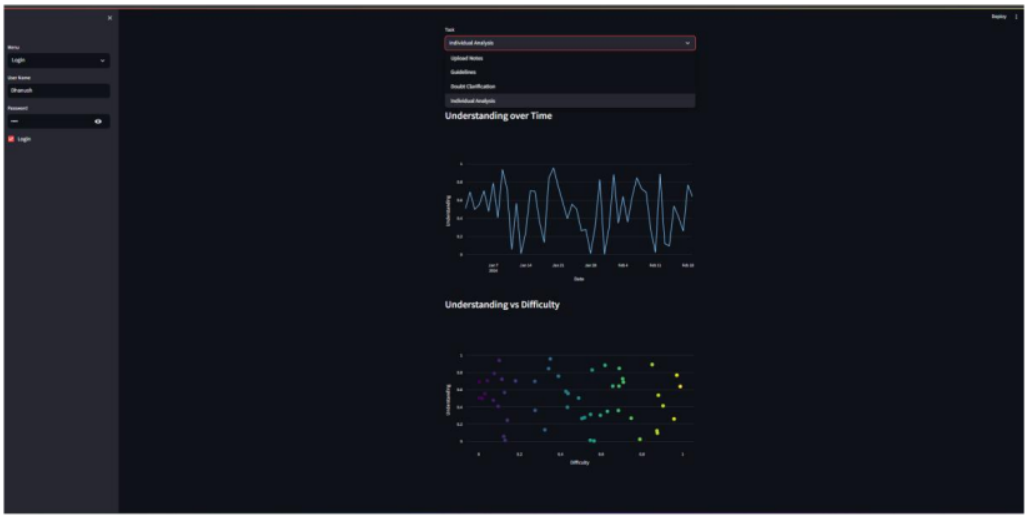

Figure 3.4.1 Student Portfolio

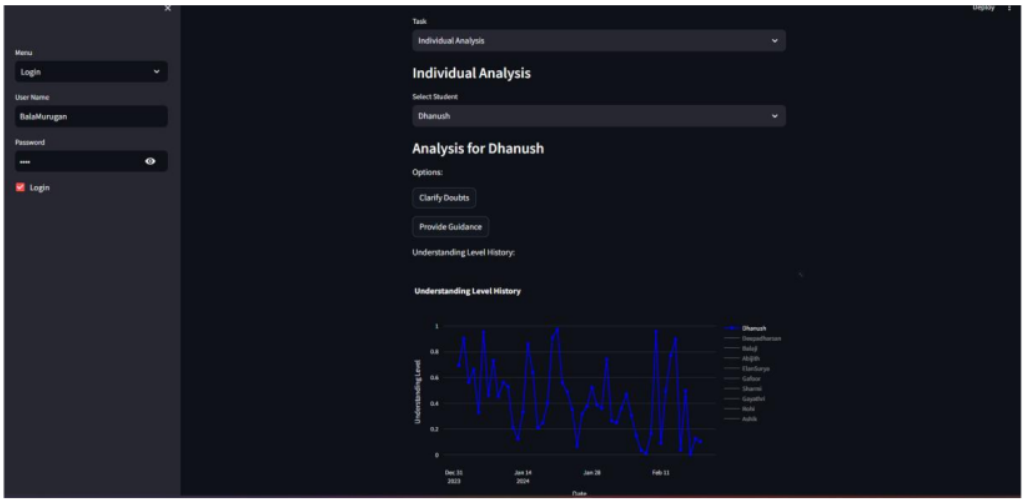

Figure 3.4.2 Staff Portfolio

IV. WORKING

In our study, we aimed to elevate our model's text comprehension and comparison skills by incorporating a diverse range of content. To achieve this, we employed web scraping to gather text data from various websites, ensuring our model was exposed to a variety of writing styles and topics. This approach enabled us to construct a robust training dataset imbued with different perspectives and writing nuances, enhancing our model's adaptability and capacity to process various text inputs creatively.

To enrich our model's training data and enhance its performance, we integrated cutting-edge AI-based models like GPT and BERT. These models allowed us to generate diverse writing styles for identical content, imparting a broader understanding of language and context to our model. By incorporating these AI models, we sought to improve our model's text recognition and generation capabilities, resembling human writing styles more closely and enhancing its effectiveness in analysing and comprehending text data.

Following the collection and preparation of the data, we meticulously pre-processed it to cleanse and standardize the text. This involved removing extraneous information such as stop words and special characters using regex, as well as identifying and eliminating anomalies. This preprocessing step ensured that our model was trained on high-quality, standardized text, which is critical for precise analysis and evaluation. Subsequently, we fine-tuned the collected data and trained it using the BERT and BoW models. BERT aided us in identifying text patterns, enabling us to calculate similarity scores between different text sets. This facilitated our assessment of the similarity between various documents, offering valuable insights into the nature of the text and enhancing our model's ability to understand and compare text data effectively.

Additionally, the BoW model played a pivotal role in our research by utilizing cosine similarity to identify overlapping words between paragraphs. This method assisted us in detecting plagiarism and ensuring the integrity of the text data used in our research, further enhancing the reliability and accuracy of our model's analysis. In summary, our study employed a comprehensive approach that integrated web scraping, AI models, preprocessing techniques, and advanced models like BERT and BoW to enhance our model's ability to recognize and evaluate text similarity. By amalgamating these techniques, we aimed to elevate the quality and effectiveness of our model in analysing and understanding text data, particularly in the realm of online learning.

Figure 4.1 illustrates the procedural steps for analyzing student engagement in a virtual classroom environment through Natural Language Processing (NLP). This workflow delineates the process of leveraging NLP techniques to assess and understand the levels of student engagement during virtual class sessions.

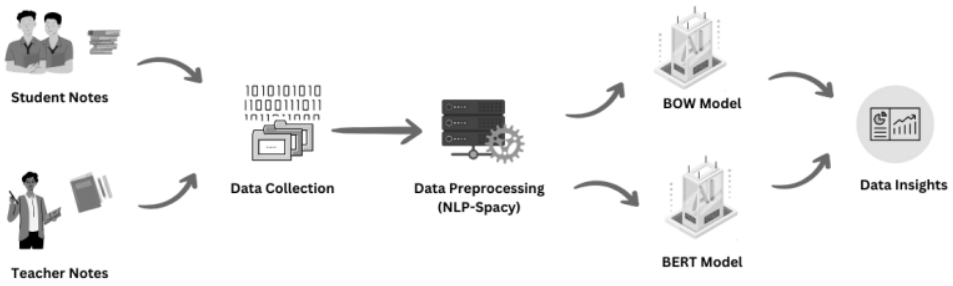

Figure 4.1 Workflow of Student Engagement Analysis in Virtual Class Using Natural Language Processing

V. EXPERIMENTAL RESULTS

The model underwent training for a total of ten epochs, with each epoch processing a batch size of thirty-two. During training, the binary crossentropy loss function was utilized, alongside accuracy as the primary evaluation metric. Progress updates were displayed throughout the training process with a verbosity level set to 1. Upon completion of the training phase, the model underwent evaluation using a separate test dataset. The test results revealed a loss value of 0.51 and an accuracy rate of 75%. The performance evaluation showcased a test loss of 51% alongside a corresponding test accuracy of 75%, providing a foundational assessment of the model's capabilities.

VI. CONCLUSION

In summary, our research has showcased the power of combining innovative techniques to enrich text analysis. By blending web scraping with advanced AI models like GPT and BERT, as well as employing meticulous preprocessing, we have significantly boosted our model's capacity to grasp and compare text. Through web scraping, we curated a diverse array of content, ensuring our model learned from a broad spectrum of writing styles and topics. The integration of AI models such as GPT and BERT enabled us to generate a myriad of writing styles for the same content, deepening our model's comprehension of language and context. Preprocessing techniques further refined the text data, guaranteeing our model trained on top-quality, standardized text.

Our approach of employing BERT for pattern recognition and BoW for plagiarism detection proved particularly effective in scrutinizing text similarity. BERT's prowess in identifying underlying patterns, coupled with BoW's knack for pinpointing plagiarism by comparing words between paragraphs, provided insightful glimpses into text similarity and integrity. These strategies significantly heightened the reliability and precision of our model's analyses. Our study underscores the transformative potential of integrating these methodologies in text analysis, especially in the realm of online learning. By enhancing our model's ability to discern and assess text similarity, we aspire to pave the way for more sophisticated and efficient text analysis tools across various domains.

VII. FUTURE WORKS

In future developments, while we aim to enhance our portfolio platform to better support students in tracking their progress and engagement in virtual classes, we anticipate certain limitations. Specifically, while our recommendation engine will utilize student notes to suggest relevant study materials, its effectiveness may be constrained by the availability and quality of the input data. Additionally, our AI-powered system to simplify lecture content may face challenges in accurately summarizing complex concepts, particularly in mathematical and language-specific contexts. Moreover, although we intend to create an online repository for lecture content to facilitate self-paced learning and revision, the accessibility and comprehensiveness of the materials may vary. Lastly, while our AI-powered doubt resolution bot aims to provide simplified explanations for complex topics, its efficacy may be limited in addressing nuanced queries or highly technical subjects. Despite these potential limitations, our overarching goal remains to enhance the virtual learning experience by offering personalized support and resources to students.

Student Engagement Analysis in Virtual Class Using Natural Language Processing

ORIGINALITY REPORT

4%
SIMILARITY INDEX

PRIMARY SOURCES

1	Aminul Islam. "Semantic text similarity using corpus-based word similarity and string similarity", ACM Transactions on Knowledge Discovery from Data, 07/01/2008 Crossref	21 words — 1%
2	summit.sfu.ca Internet	21 words — 1%
3	Yuanyi Zhen, Jar-Der Luo, Hui Chen. "Prediction of Academic Performance of Students in Online Live Classroom Interactions—An Analysis Using Natural Language Processing and Deep Learning Methods", Journal of Social Computing, 2023 Crossref	18 words — 1%
4	doaj.org Internet	15 words — 1%
5	www.denverhealth.org Internet	12 words — < 1%
6	Yeji Song, Jihwan Shin, Jinhyun Ahn, Taewhi Lee, Dong-Hyuk Im. "Generating Labeled Multiple Attribute Trajectory Data with Selective Partial Anonymization based on Exceptional Conditional Generative Adversarial Network", IEEE Access, 2023 Crossref	10 words — < 1%

EXCLUDE QUOTES	OFF	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	OFF	EXCLUDE MATCHES	OFF