A Study on the Perception of Hybrid Education

Dr.P. Aruna[*1], Dr. D.Manju[*2], Swetha Srinivasan[*3], Akkshaya Sri J[*4], Hari Priya H[*5], Sanchez Innocencia D[*6]

[*1*2] Assistant Professor, Department of Computing, Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India

[*3*4*5*6] M. Sc (Integrated) Decision and Computing Sciences, Department of Computing, Coimbatore Institute of Technology, Coimbatore, Tamil Nadu, India

Abstract:

COVID-19, a random outbreak, had a serious effect on every living form and changed the education system dramatically. COVID-19 had a huge impact on both students and teacher's lifestyle especially by affecting the environment, mental and physical health. It has changed the perspective of teaching for students as well as the teachers. Our research aims to perform a profound study on the influence of COVID-19 from the perspectives of teachers and students. This study aids in developing a solid strategy to upgrade the existing education system to provide standard education and also tackle potential difficulties in students and teachers' perspective on their education and career which aims to analyse, interpret and drive out a decision.

Keywords: COVID-19, hybrid education, prediction, online education

I Introduction

Opinions on the internet are vast and rising day by day across the globe on the platforms such as online blogs, social platforms like facebook, twitter, linkedin, quora etc. Users do not limit their opinions on products or services, they literally speak their heart out on the things that they are either directly or indirectly affected by. Among the different sources that are available of opinionated data, quora is an online platform where even students, parents and teachers everyone use it for their own purposes of learning, reading and it would be an apt fit to extract the data from it as it depends on the different perceptions of the people who are directly linked to it. As the study depends on different users and opinions, this study falls under the field of Natural Language Processing.

This study aims to extract the opinions of the users from their comments or feedback that they have shared on the online source. With ever creeping opinions on the internet, this study focuses on scraping out the comments and feedback related to the study after the pandemic. This is done by web scraping of text which is a field of Natural Language Processing and it specially focuses on capturing user opinions by their provided context. The obtained textual data are segmented using the topic modelling method. The data collected takes a big part in the study as it aids to distinguish the individual's opinion, concern and how they are directly or indirectly affected. The tool that is developed is an evaluation tool for a vast collection of opinions on quora. By this model, it becomes easier to analyse the data of opinions in a broad way. The main purpose of the paper are

- To collect the different opinions.
- To develop and summarise the opinion of the users using topic modelling.
- To evaluate how the factors vary according to different individuals and how their perception changes on hybrid education.

II Literature Review

Bhatia et. al. in their paper titled An Improved Method for Extractive Based Opinion Summarization Using Opinion Mining they have developed the need of automatic summarization efficiently resulting in increased interest among communities of Natural Language Processing and Text Mining. This paper emphasis on building an extractive summarization system combining the features of principal component analysis for dimensionality reduction and bidirectional Recurrent Neural Networks and Long Short-Term Memory (RNN-LSTM) deep learning model for short and exact synopsis using seq2seq model. It presents a paradigm shift with regard to the way extractive summaries are generated. Experiments on the different datasets also outperforms the previous researches and the accuracy is claimed to achieve more than the baselines, showing the efficiency and the novelty in their research paper.

Hariharan, S., et. al., in their paper titled, "Opinion mining and summarization of reviews in web forums" focuses on providing a methodology for mining the opinions using generic user focused reviews. In their paper, they proposed a technique that extracts the opinion words from the reviews. By using the extraction algorithm they have assigned scores to each of the words in the review. Based on the cumulative weight obtained the opinion miner decides

whether to recommend the product to the user or not. The method is domain independent and can be applied to any domain. Moreover they have said that this method can be applied to any review format, provided the reviews are structured and formatted.

F. Es-Sabery et al., in their paper titled, "A MapReduce Opinion Mining for COVID-19-Related Tweets Classification Using Enhanced ID3 Decision Tree Classifier" focuses on Opinion Mining (OM) which aims to capture human sentiment in the given text. The textual data (reviews, tweets, or blogs) was classified into three different class labels which are negative, neutral and positive for analysing and extracting relevant information. In this contribution, they introduced an innovative MapReduce improved weighted ID3 decision tree classification approach for OM. Their work is implemented in a distributed environment using the Hadoop framework, with its programming framework MapReduce and its distributed file system HDFS. Their primary goal is to enhance the performance of a well-known ID3 classifier in terms of accuracy, execution time, and ability to handle the massive datasets. They have carried out several experiments that aim to assess the effectiveness of our suggested classifier compared to some other contributions chosen from the literature. The experimental results demonstrated that their ID3 classifier works better on COVID-19.

Gupta, I., et.al., in their paper titled "Enhanced Twitter Sentiment Analysis Using Hybrid Approach and by Accounting Local Contextual Semantic" has addressed the problem of Twitter sentiment analysis through a hybrid approach in which SentiWordNet (SWN)-based feature vector acts as input to the classification model Support Vector Machine. Their main focus is to handle lexical modifier negation during SWN score calculation for the improvement of classification performance. And they present a naive and novel shift approach in which negation acts as both sentiment-bearing word and modifier, and then they shift the score of words from SWN based on their contextual semantic, inferred from neighbouring words. Experimental results had shown that the contextual-based SWN feature vector obtained through shift polarity approach led to an improved Twitter sentiment analysis system that outperforms the traditional reverse polarity approach. They validated the effectiveness of the hybrid approach considering negation on benchmark Twitter corpus.

Kumar, A., et. al., in their paper titled "Opinion Extraction from Quora Using User-Biased Sentiment Analysis" has presented an opinion extraction model based on the user's profile. The opinion formulation algorithm was governed by factors that vary among users. A user-biased sentimental analysis technique was introduced, which mines the answers written on various topics on the popular Web site Quora and provides an opinion based on the user's preferences. For implementation, a personal assistant to assist students in selecting a university for graduate studies based on their preferences, of course, the return of investment expectations, importance to university ranks, etc was created. The algorithm achieved optimal performance and hence can be used as a reliable method to form opinions on behalf of the user.

G Elangovan, et.al., in their paper titled "Medical Quora Tagging using MATAR and LDA Algorithm", has developed an enhanced Latent Dirichlet Assignment clustering method & Inter Modeling for Tag Suggestion rating system is documented to boost correlation - based & identification efficiency to suggest labels with material modern web labels that promotes the exchange of medical information using unmonitored data through question-answering. For accurate tagging, Methods like POS marking, Hopping, Whistles& Stopping words are being used for speech recognition. The efficiency of the evolved architectures is compared to the

standard methods, by using specificity of the recommendation, defining features, sensitivity, plain word and speed. Their findings have revealed that the classification and grouping scheme of the proposed structure succeeds traditional textual record approaches.

Nwadiugwu MC and Nwadiugwu CC(2021) in their paper titled “Influencing Followership: Understanding the Perspective of those Leading Active Discussions on Quora" presented a qualitative study of eight participants leading active discussions on Quora was conducted using semi-structured in-depth interviews, followed by thematic analysis. The open coding method was used to iteratively code related answers to develop themes. Results suggest that copyright tactics, controversial answers and sharing new information are some of the mechanisms for influencing followership. These mechanisms are built overtime through conscious strong engagement and by writing a consistently well-thought-out answer. The motivation for leading and writing answers on Quora were more intrinsic than extrinsic, and most participants believed influencing followership should not be a concern if one has the right message.

III Methodology

3.a. Data Cleaning and Pre-processing:

Data aggregated using the tool contains a lot of noise which affects the modelling process. The data needs to be cleaned before being pre-processed to achieve accurate results. Microsoft Excel’s feature of filters was employed to clear the noise from the data.

Data pre-processing is a step in which the data gets transformed, or encoded, to bring it to such a state that the machine can easily parse it. Text data contains noise in various forms like emotions, punctuation, text in a different case, which need special type of pre-processing to make it ready for the model. The type of pre-processing also depends on the domain or business involved. The following section describes the various methods of pre-processing techniques carried out on the data before being modelled.

3.b. Tokenization, Lowercasing, Removal of Punctuations and Digits:

The text data is split into tokens to represent them in a vector space model. The tokens are further converted into lower case to avoid multiple dimensions in the vector space model for the same word. Numbers and punctuation marks are further removed to avoid the results being affected.

3.c. Stopwords Removal:

Stopwords are words that occur frequently and add no value to the analysis carried out. These words include the articles, misspelt words, etc. Since they do not carry much information, those words can be eliminated to focus on more important features. Apart from this, subject-wise stopwords also need to be removed with the help of domain knowledge. Along with stopwords available with the NLTK library in Python, certain domain based stopwords were also removed from the corpus to enhance the performance of the model.

3.d. Lemmatization:

Lemmatization is one of the most common text pre-processing techniques used in Natural Language Processing (NLP) to reduce a given word to its root word which is known as 'lemma'. Lemmatization is chosen as it is more accurate than stemming because of its use of algorithms and knowledge to provide words with dictionary meaning.

3.e. Creation of Bigrams and Trigrams:

A bigram or digram is a sequence of two adjacent elements from a string of tokens while a trigram is a sequence of three adjacent elements. Identifying and creating bigrams and trigrams help comprehend the data in an effective manner by providing better contextual understanding. Gensim, a package in Python library dedicated for topic modelling provides a feature named 'phrases', to help encode bigrams and trigrams.

3.f. Data Modelling

The pre-processed data is further fit to a Natural Language Processing (NLP) model, Latent Dirichlet Allocation (LDA) to identify various themes present in the conversations.

3.g. Topic Modelling

In machine learning and natural language processing, topic models are generative models, which provide a probabilistic framework. Topic modelling methods are generally used for automatically organising, understanding, searching, and summarising large electronic archives. The "topics" signifies the hidden, to be estimated, variable relations that link words in a vocabulary and their occurrence in documents. A document is seen as a mixture of topics. Topic models discover the hidden themes throughout the collection and annotate the documents according to those themes. Each word is seen as drawn from one of those topics. Finally, a document coverage distribution of topics is generated and it provides a new way to explore the data on the perspective of topics.

3.h. Latent Dirichlet Allocation

Latent Dirichlet allocation model (LDA) is a generative probabilistic topic model where each document is represented as a random mixture of latent topics and each topic is represented as a distribution over a fixed set of words. LDA aims to identify the underlying latent topic structure based on the observed data. In LDA, the words of each document are the observed data. For each document in the corpus, the words are generated in a two-staged procedure. First, a distribution over topics is randomly chosen. Based on this distribution, a topic from the distribution over topics is randomly chosen for each word of the document. In LDA, a word is a discrete data from a vocabulary indexed by {1, …, V}, a sequence of N words w=(w1, w2, …, wn) and a corpus is a collection of M documents denoted by D={w1, w2,…, wM}.

The process of LDA can be modelled by a three-level Bayesian graphical model, where random variables are represented by nodes and possible dependencies between the variables

are represented by edges, as depicted in Figure 3.1. In this representation, α refers to Dirichlet parameter, Θ refers to document-level topic variables, z refers to per-word topic assignment, w refers to the observed word and β refers to the topics. As it can be observed from the three-layered representation, α and β parameters are sampled once while generating the corpus, document-level topic variables are sampled for each document and word-level variables are sampled for each word of the document.

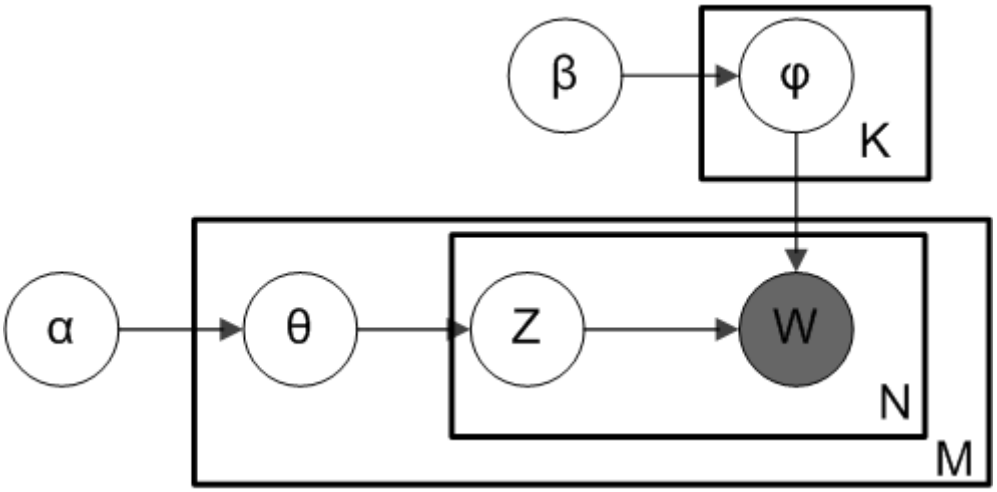

Figure 1: Graphical representation of LDA Model

3.2. Model Optimization:

Performance Measures:

The effectiveness of topic classification in LDA models is closely related to the selection of the number of topics. The optimization process involves choosing the optimal number of topics as the model itself does not provide the optimal number.

Topic Coherence:

Topic coherence measures the consistency of a single topic by measuring the semantic similarity between words with high scores in a topic, which contributes to improving the semantic understanding of the topic. That is, words are represented as vectors by the word's co-occurrence relation, and semantic similarity is the cosine similarity between word vectors. The coherence is the arithmetic mean of these similarities. The Coherence Model from Gensim (RARE Technologies Ltd), the Python package for natural language processing, to calculate the coherence value. To compute topic coherence of a topic model, the following steps are performed:

1. Select the top n frequently occurring words in each topic
2. Compute pairwise scores (UCI or UMass) for each of the words selected above and aggregate all the pairwise scores to calculate the coherence score for a particular
   $Coherence = \Sigma_{i<j} score(w_i, w_j)$

3. Take a mean of the coherence score per topic for all topics in the model to arrive at a score for the topic model.

Once the coherence scores are calculated for each value of k, (number of topics), the values of k and their corresponding coherence scores are plotted on a graph to identify the optimal number of topics. The value of k with the highest coherence score gives us the optimal number of topics. When the value of k is too large, there are high chances of the same keywords being repeated in multiple topics.

Perplexity:

The perplexity method considers the predictive ability of the model for documents. The principle is that if most of the words in a document are the top words in the probability ranking of the topics generated by LDA, then the prediction of the topic to which the document belongs is more accurate. On the contrary, it is difficult to determine the topic of the document. Therefore, the lower the perplexity, the better the prediction ability of the model. Perplexity is calculated as follows:

$$perplexity(D_{test})=\exp\{-\Sigma^{M}_{d=1}\log p(w_d)\Sigma NdMd=1\}$$

Where M is the size of the corpus,
N_d is the size of the d^{th} text (number of words) and
$p(w_d)$ denotes the probability that the word w_d is generated in the document d.

Existing studies found that although the perplexity can judge the predictive ability of the topic training model to a certain extent, when the number of topics is selected by the perplexity, the number of topics selected is often too large and does not converge, and similar topics are likely to appear, resulting in a low discrimination of topics. As depicted in Figure 2, value of k =15 was chosen as optimal value based on the graph.

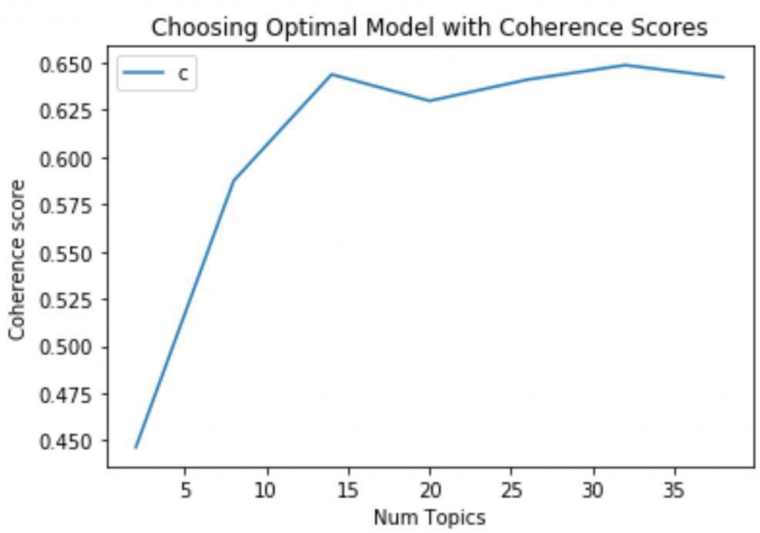

Figure 2: Coherence score graph

IV RESULTS DISCUSSION

Figure 3: Perspective of Hybrid Course

Figure 2 shows that people are talking about the self course, website, lecture, application for a hybrid study. It is considered to be well advantageous to the students who work on engines, battery as they might get a clear idea on the demonstration, inner workings and the process flow of the engines by viewing videos on the website and they also get a hands on experience on the machines when they go to their respective educational institutions.

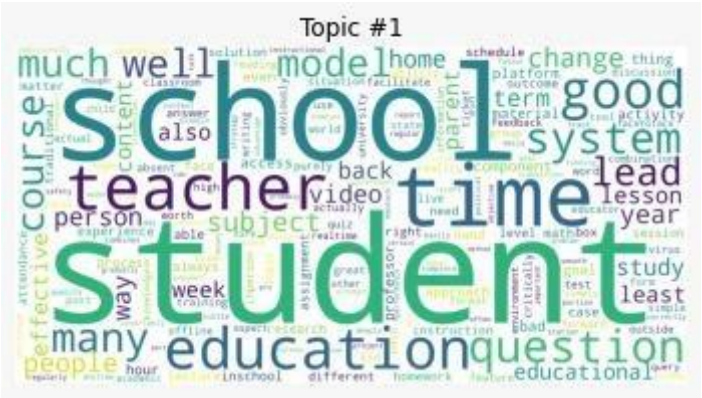

Figure 4: Perspective of School Students on Hybrid Education

Figure 3 depicts that there are many ways for the students to get their education and one of the best ways is opting for hybrid education. When the students opted for hybrid education, they came up with many educational questions on the lesson and course. The students also learnt to manage time and they feel the hybrid education system is a good system as they get different opportunities to explore their lessons, courses rather than listening to the lectures whole week.

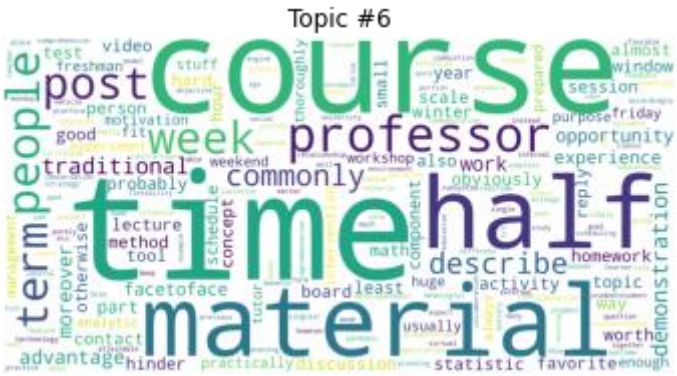

Figure 5: Perspective of Professors on Hybrid Education

The above figure depicts the opinion of professors regarding blended learning which is positive. The professors feel that it saves a lot of time while teaching in online mode. The material for the course can be enriched in the time saved while having access to various sources of information from the internet. The professors also feel they have enough time and opportunities to pursue online courses to enhance their knowledge.

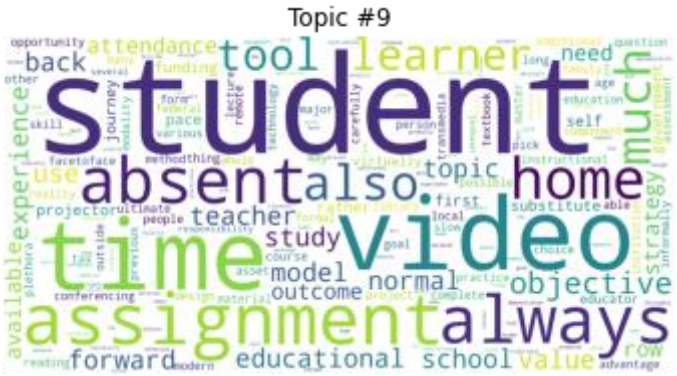

Figure 5: Perspective of Students on Assignments in Hybrid Learning

The absentee rate of students was low on the online educational mode as it was easier for them to attend from anywhere on the devices with good internet connection. The objective of this type of educational mode is to make them learn with a deep understanding of the subjects they go through. The video learning has made a good impact on the students as they tried to explore new tools for the models they use in their project. Though they felt like there were many assignments given to them, there is still a study telling that if the assignments are interesting and challenging, they take it on a positive aspect. This somehow increases the school value too.

V. CONCLUSION

In hybrid education, learning becomes more liberal, adaptable and interesting for the students and to the institutions. Based on the results found, the key trait of hybrid education is that it can be quickly modified and regulated as per the requests and needs of the learner. The course and some of the few measures such as time management and assignments tell how the students are positively managing and are looking forward to learning new things. With ease of learning from anywhere and at any time, the professors and students are able to quickly adapt to new

situations. Apart from school and college students, normal people can also make use of this approach to kindle their curiosity and gain knowledge to improve their passion. This mode of education caters to the needs of all types of interested learners. Hybrid education paves the way for the institutions to deliver their education in a contemporary way via both in-person and online mode. Time spent at educational institutes will still be critical since it will allow the students to improvise their practical knowledge and skills and engage on-the-ground to enhance their social skills that will majorly contribute to a comprehensive hybrid learning experience. Hybrid education paves the way for the institutions to deliver their education in a contemporary way via both in-person and online mode.

References:

1. S. Bhatia and M. AlOjail, "*An improved method for extractive based opinion summarization using opinion mining*," Computer Systems Science and Engineering, vol. 42, no.2, pp. 779–794, 2022.
2. Es-Sabery, F., Es-Sabery, K., Qadir, J., Sainz-De-Abajo, B., Hair, A., Garcia-Zapirain, B., & De La Torre-Diez, I. (2021). "*A MapReduce Opinion Mining for COVID-19-Related Tweets Classification Using Enhanced ID3 Decision Tree Classifier*". IEEE Access, 9, 58706–58739.
3. Nwadiugwu MC and Nwadiugwu CC(2021) "*Influencing Followership: Understanding the Perspective of those Leading Active Discussions on Quora*".Front. Computer. Sci. 3:582242.
4. G Elangovan, J Umamageswaran, G Indumathi and A V Kalpana(2021), "*Medical Quora Tagging using MATAR and LDA Algorithm*", Journal of Physics: Conference Series, Volume 1964, Advances in Computer Science Engineering.
5. Gupta, I., & Joshi, N. (2019), "*Enhanced Twitter Sentiment Analysis Using Hybrid Approach and by Accounting Local Contextual Semantic*", Journal of Intelligent Systems, 29(1): 1611–1625.
6. Kumar, A., Praveen, S., Goel, N., & Sanwal, K. (2018). "*Opinion Extraction from Quora Using User-Biased Sentiment Analysis*", Information Systems Design and Intelligent Applications, 219–228.
7. Hariharan, S., Srimathi, R., Sivasubramanian, M., & Pavithra, S.(2010), "*Opinion mining and summarization of reviews in web forums*" Proceedings of the Third Annual ACM Bangalore Conference on - COMPUTE '10.