Simplifying the Process of Learning with the Help of Machine Learning and Natural Language Processing

Priyanka Patel

ABSTRACT
The difficulty of sifting through the multitude of internet resources is addressed by Easyfy, a software program created to streamline the learning process. Easyfy seeks to help students find the best resources in the age of readily available internet content by applying a set of rules or criteria. Since popularity or view count by themselves may not always imply the quality of instructional content, Easyfy's main focus is on YouTube resources. Easyfy, on the other hand, takes a novel tack by scrutinizing user commentary on YouTube videos. Easyfy distinguishes between good and negative feedback by analyzing and interpreting the comments to determine the sentiment conveyed. Sentiment analysis of YouTube comments is essential for gauging a video's overall popularity. Easyfy helps people comprehend the ratio of positive to negative sentiment by measuring the sentiment of comments. This insightful information can help students make well-informed decisions about the suitability and connection of a certain video to their learning goals. Sentiment analysis in YouTube comments is used for purposes other than personal preference. Easyfy uses this data to suggest the "best" video for a particular subject. The program analyzes the tone of comments in addition to popularity measures to identify which video is most likely to offer the most thorough and well-received information on a certain topic.

Keywords
Machine Learning, Easy learning, feedback, scrutinizing

1. INTRODUCTION
Easyfy is a complete platform that aims to transform the field of educational support. In the modern day, with the internet serving as the primary means of obtaining knowledge, Easyfy appears as a beacon of hope for those in need of assistance with their academic pursuits. Understanding the enormous amount of information available online, especially the abundance of instructional videos on sites like YouTube, Easyfy aims to make learning easier by offering a customized solution for effective video finding and consumption. Easyfy understands the importance of using the internet, especially YouTube, as a primary learning medium in the academic setting, where time is of the essence. The site meets the demands of visitors who want to learn more about various disciplines, especially in the field of mathematics. Finding the most relevant and useful videos among the multitude of possibilities on YouTube presents a difficulty as we navigate the vast ocean of content that is readily available. This is where Easyfy, a user-centric tool, comes in. By offering personalized recommendations based on the user's search query, it relieves the stress of sorting through a large number of films. Imagine a situation when a student wants to understand a specific mathematical idea. Easyfy makes it easier to find the most insightful and educational videos on a given topic by compiling a list of the best ones, saving you the trouble of having to trawl through a ton of videos. Easyfy recognizes that users and students have limited time, thus it works to reduce the overwhelming amount of content available and allow users to concentrate on the most useful learning resources. The core of Easyfy is its capacity to provide users with a tailored and effective answer to their educational questions. Easyfy uses cutting-edge algorithms and machine learning to find related movies and then refines the list to show you the best possible alternatives. By enabling users to make well-informed decisions on their learning path, this strategic curation helps them save valuable time and promotes a more productive learning environment. Easyfy is really more than a tool; it is an educational partner that is aware of the changing nature of contemporary learning. Easyfy serves as an invaluable resource for students, learners, and anyone looking to expand their knowledge base by providing the most relevant and impactful educational content available on platforms like YouTube. It does this by bridging the gap between the abundance of online content and the need for focused, efficient learning.

2. LITERATURE REVIEW
The literature review serves as a critical examination of existing research in the field of [Your Topic]. This comprehensive review aims to provide a foundation for the current study, offering insights into the current state of knowledge, identifying gaps, and contextualizing the research within the broader academic discourse. Wankhade, M., Rao, A.C.S. and Kulkarni, C. has explained that The swift expansion of web-based programs, such blogs and social media sites, has led to remarks and evaluations of daily activities. The process of obtaining and examining people's views, ideas, and perceptions about various subjects, products, and services is known as sentiment analysis. Corporations, governments, and individuals can all benefit from people's opinions when gathering data and making opinion-based decisions[1]. With the use of specially designed extraction tools, the writers retrieved 29,386 comments from 150 instructive YouTube videos. Analysis of sentiment and qualitative content was done.This study was done by CS Lee, H Osop, DHL Goh, G Kelni[5]. A research study by H Bhuiyan, J Ara, R Bardhan, MR Islam in 2017 also spread light on this topic[7]. In 2019, S Nawaz, M Rizwan, and M Rafiq proposed an innovative approach utilizing Google API for quantitative sentiment analysis of YouTube video comments, yielding a single normalized score to recommend videos as "not recommended," "may be recommended," "recommended," or "highly recommended," enhancing video selection criteria[8]. In 2017, J Savigny and A Purwarianti conducted a study on emotion classification in Indonesian YouTube comments, utilizing a corpus of 8,115 comments manually labeled for six basic emotions and one neutral label. The study compared various word embedding methods, including average word vector, average word vector with TF-IDF, paragraph vector, and Convolutional Neural Network (CNN). The CNN method demonstrated the best performance, achieving an accuracy of 76.2%, surpassing the baseline SVM with Unigram TF-IDF[11]. In a study by M Aufar, R Andreswari, and D Pramesti, sentiment analysis on public comments about Nokia's products on YouTube was conducted using Decision Tree and Random Forest algorithms. Despite neutral sentiments dominating, the Decision Tree algorithm

achieved a slightly higher accuracy of 89.4% compared to Random Forest's 88.2%, offering insights into the overall perception of Nokia's product quality based on social media opinions[16]. In their study, A Patil and S Gupta address the challenges of sentiment analysis in the context of widespread social media usage. The research highlights the limitations of existing machine learning and lexicon-based techniques and advocates for the development of new automated methods to achieve improved accuracy and overcome challenges in sentiment analysis[3]. In their study, GS Chauhan and YK Meena explore the underutilized potential of user comments on YouTube in the context of video retrieval and ranking. Emphasizing the limitations of current rating-based approaches, the research employs aspect-based sentiment analysis to investigate the impact of various aspects of a video's subject as expressed in user comments, aiming to enhance the relevancy and quality determination in video ranking[20]. Conclusively, the vast array of examined literature highlights the increasing importance of sentiment analysis and content ranking in the context of user-generated material on websites such as YouTube. Scholars, such as the aforementioned individuals, have explored novel approaches like aspect-based sentiment analysis, which has transformed our comprehension of how viewpoints and interactions impact the relevancy and caliber of material. This paper highlights the necessity for continued research to investigate various ways as the digital landscape continues to change, offering a thorough grasp of sentiment analysis and content ranking in the big data and Internet of Things era. Because this discipline is interdisciplinary, it encourages ongoing research and adaptation to take advantage of the opportunities and problems that user-generated material on various online platforms presents.

3. METHODOLOGY

The data collection process for YouTube comments relevant to educational content follows a systematic approach, ensuring diversity and comprehensive representation.

3.1 Data Collection

3.1.1 Define the Scope

Clearly delineate the scope for sentiment analysis, identifying specific educational topics or content categories, ranging from academic subjects to tutorials.

3.1.2 YouTube API Access (Optional)

Consider leveraging the YouTube API for efficient, programmatic access to comments from designated videos or channels, facilitating the retrieval of a substantial comment volume.

3.1.3 Manual Collection

Alternatively, manually curate YouTube comments by selecting a diverse set of videos within the defined scope, covering various aspects of chosen educational topics.

3.1.4 Scrapping Tools (Caution)

Exercise caution with web scraping tools, ensuring compliance with YouTube's terms of service, and being mindful of potential legal and ethical implications.

3.1.5 Diverse Representation

Strive for diversity in the selected videos to create a representative dataset, encompassing varying view counts, likes, and dislikes, capturing a spectrum of perspectives.

3.1.6 Data Privacy Considerations

Respect user privacy, adhere to data protection regulations, and avoid collecting personally identifiable information (PII), maintaining ethical standards in handling user-generated content.

3.1.7 Sampling Strategy

Define a sampling strategy for a balanced representation of comments, considering both popular and less popular videos, spanning different time periods.

3.1.8 Comment Metadata

Capture additional metadata for each comment, including the commenter's username, timestamp, and relevant contextual information, offering insights into user engagement patterns.

3.1.9 Data Storage

Store the collected data in a structured format such as a CSV file or database, ensuring a sufficient number of comments for meaningful sentiment analysis and facilitating ease of analysis. This meticulous approach to data collection forms the foundation for robust sentiment analysis within the educational content landscape.

A wide range of educational themes were covered in the broad dataset of YouTube comments gathered for this sentiment analysis study. The dataset consists of comments from tutorials, academic subject-related films, and other instructional resources. It was noted during the data gathering procedure that the opinions conveyed in the comments differed depending on the kind. The gathered dataset's sentiment distribution looks like this: There were 572 "Negative," 1689 "Irrelevant," 2333 "Neutral," and 519 "Positive" comments. A thorough examination of user feedback is made possible by the dataset's varied representation of sentiment types, which also provides insights into the spectrum of opinions expressed in relation to YouTube instructional content. Incorporating 'Neutral' and 'Irrelevant' sentiments enhances the analysis's depth by encompassing a range of user reactions that extend beyond either affirmative or negative statements. In later stages, this distribution will be investigated and examined in more detail in order to obtain valuable insights into the sentiment landscape of the educational content that is being examined.

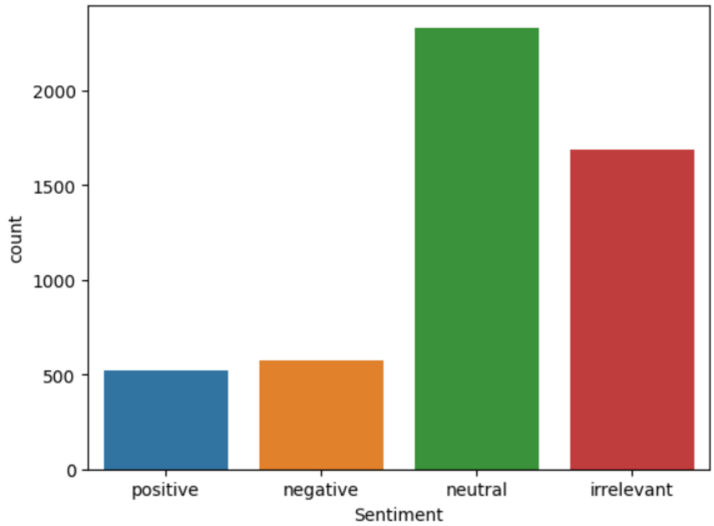

Figure 1 Sentiment Distribution in the Dataset

3.2 Data Preprocessing

3.2.1 Text Cleaning

One of the most important steps in getting the YouTube comment data ready for sentiment analysis is text cleaning. In order to maintain consistency throughout the dataset, it is important to make sure that the raw text of comments is free of superfluous characters, symbols, and URLs. Text data is standardized through the use of text cleaning techniques, which include removing special letters and symbols. This lays the groundwork for precise sentiment analysis.

3.2.2 Stopword Removal:

The goal of stopword removal is to remove frequently occurring stopwords that add nothing to the understanding of sentiment analysis in the process of fine-tuning the dataset for sentiment analysis. The procedure entails employing a predetermined set of stopwords, which comprise such like "the," "and," and "is." The focus is shifted to terms that convey sentiment by eliminating these frequently used words, improving the caliber and applicability of the study.

3.2.3 Lemmatization or Stemming:

By returning inflected words to their common base form, lemmatization and stemming aim to standardize words. By treating similar words similarly, this procedure helps achieve word normalization, which enhances sentiment analysis accuracy. Lemmatization and stemming are the two options available to researchers depending on the particular needs of their analysis.

3.2.4 Data Integrity Checks:

Maintaining the dataset's integrity is essential to getting significant sentiment analysis results. Finding and fixing any inconsistencies in the dataset is the aim of data integrity checks. This entails running checks for anomalies, duplicate comments, and missing values. The quality and dependability of the dataset are preserved throughout the analysis by taking care of these problems.

3.2.5 Tokenization:

Tokenization is the process of separating the preprocessed text into discrete words or tokens. Here, the goal is to make it easier to analyze individual words, enabling a more detailed investigation of the sentiment conveyed in each comment. The cleaned text is divided into tokens using a tokenizer, which helps with the efficient processing of the data in later phases.

3.2.6 Handling Emoji and Emoticons:

Handling sentiment-containing items, such as emoticons and emojis, is a complex part of data preprocessing. The goal is to determine if emoticons and emojis should be kept or removed based on how relevant they are to sentiment. Emojis are a useful tool for adding context to sentiment analysis and expressing complex emotions. The goals of the research and the intended level of sentiment analysis inform how these components are handled.

3.2.7 Lowercasing:

The process of standardizing text to lowercase is done to guarantee analysis consistency. Here, the goal is to convert all text to lowercase in order to minimize any possible problems that may arise from case sensitivity in sentiment analysis. By treating words consistently, this stage improves the precision of later analytical procedures..

3.2.8 Data Formatting:

Preparing the preprocessed data for efficient sentiment analysis is the aim of data formatting. This entails arranging the cleaned and processed data in a way that makes sense for the sentiment analysis algorithm or tool of choice. For smooth integration with later phases of the research and to guarantee that the dataset is best organized for precise sentiment insights, proper data formatting is crucial.

Application of Data Preprocessing

In the research, the described data preprocessing steps were implemented using Python programming language and relevant libraries. The code included functions to clean text, remove stopwords, perform lemmatization, and handle other aspects of data preprocessing. This ensured that the YouTube comments were transformed into a structured and standardized format suitable for subsequent sentiment analysis.

3.3 Sentiment Analysis

The objective of the Sentiment Analysis phase was to classify YouTube comments into four different sentiment categories: "Neutral," "Irrelevant," "Negative," and "Positive." Text vectorization was a critical step prior to the use of machine learning models. This required converting textual input into a numerical format for analysis.

In order to have a more profound understanding of the features of YouTube comments, the distribution of text lengths was represented visually across various sentiment categories. The frequency distribution of text lengths for each sentiment—positive, negative, neutral, and irrelevant—is depicted in a histogram that was created using Seaborn. The length of the processed text is shown by the x-axis, while the frequency of occurrences is shown by the y-axis. The distribution of text lengths within each sentiment category is shown in detail by the histogram's bins and kernel density estimate (kde). A more sophisticated comprehension of the potential correlation between sentiment expressions in the YouTube dataset and comment length is made possible by the discernible patterns seen in the graph. In addition to complementing the heatmap representations of sentiment analysis results, this visualization improves the overall investigation of sentiment characteristics.

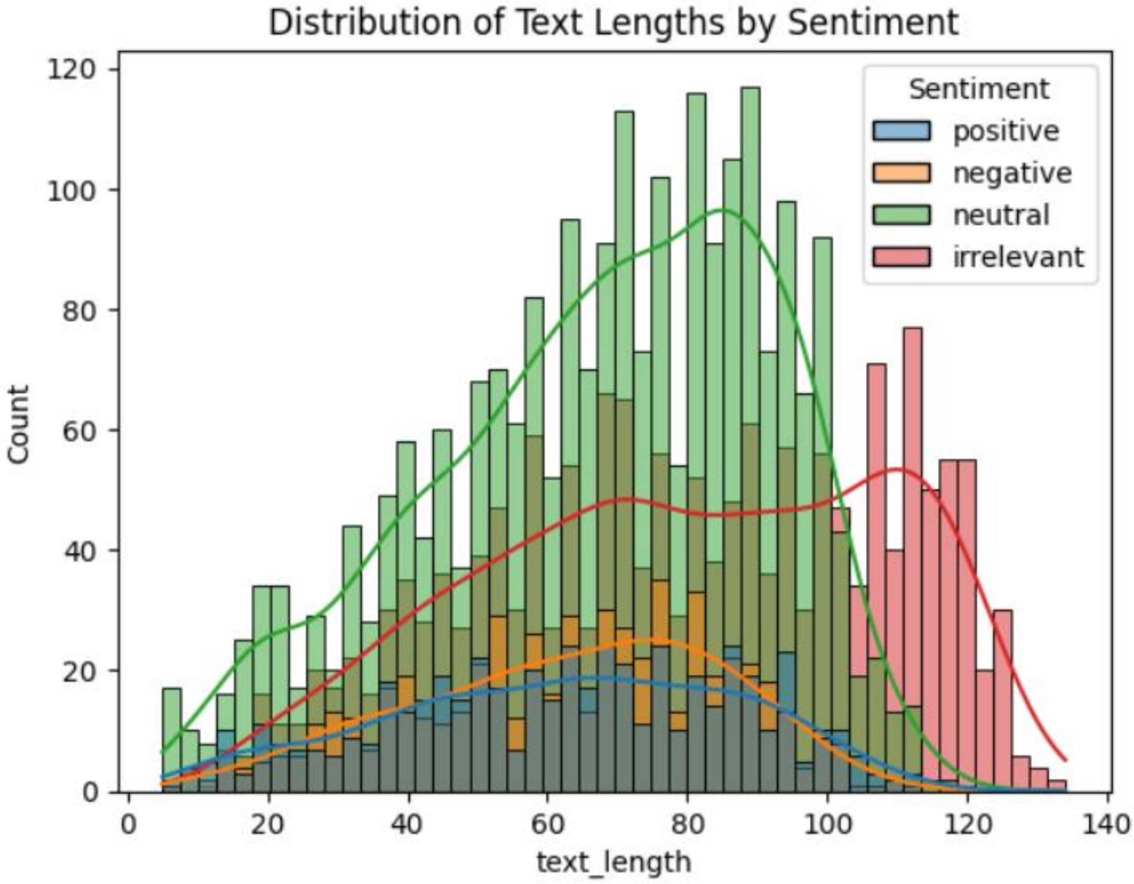

Figure 2 Distribution of Text length by sentiment

A word cloud was created to provide a visually appealing depiction of the most common words linked to positive sentiment. The core of positive phrases found in the YouTube comment collection is captured by the positive sentiment word cloud. Words are enlarged in this image according to how frequently they occur, giving users an instant overview of the popular keywords linked to favorable feelings. The more bold and larger the word, the more often it appears in positive remarks. The generated word cloud is a useful and dynamic resource for comprehending the terms and ideas that are frequently linked to good feelings in the YouTube comments.

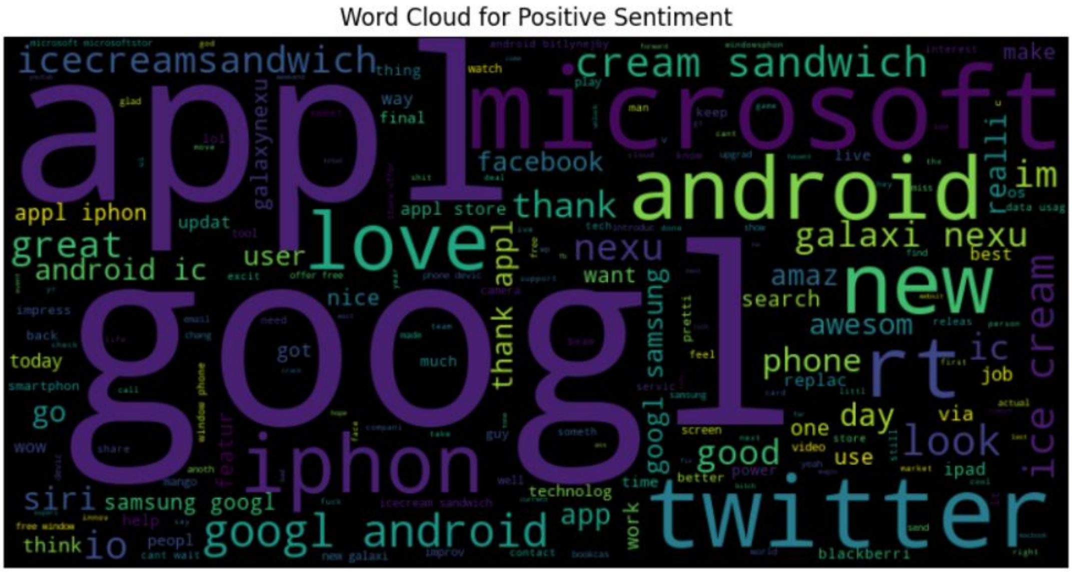

Figure 3 word cloud for positive comments

In a similar vein, a word cloud was created to display the most common terms in comments classified as derogatory emotion. The vocabulary and reoccurring motifs linked to the unfavorable sentiments conveyed in the YouTube comments are strikingly represented by this image. Words are ranked by frequency, making it possible to quickly determine the main terms that contribute to the category of negative sentiment. The generated word cloud makes it easier to spot the linguistic motifs and attitudes that might be present in remarks that have a critical tone.

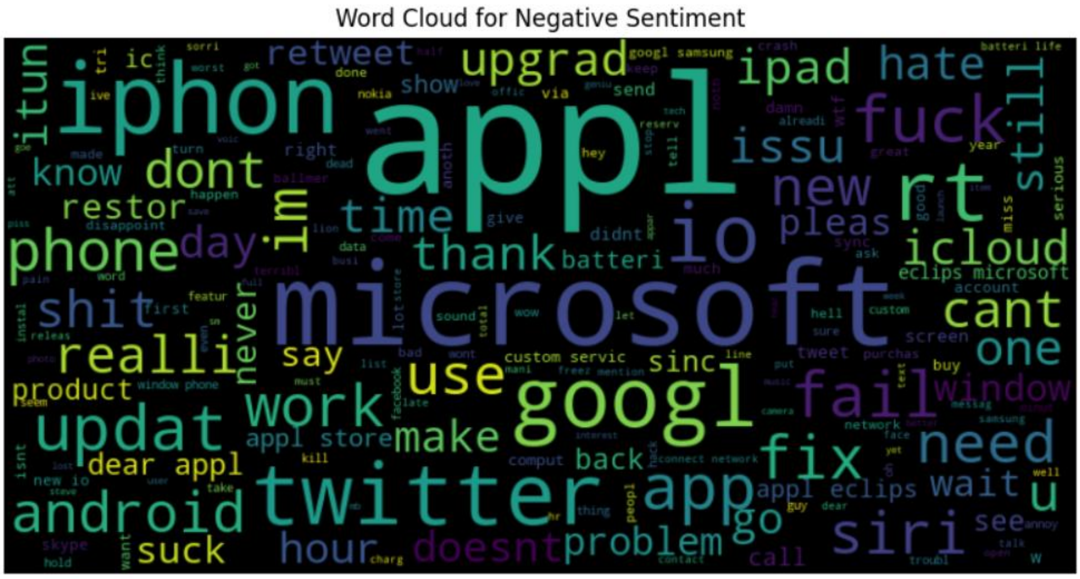

Figure 4 Word Cloud for Negative Sentiment

In this study, two vectorization methods were used: Count Vectorization and TF-IDF Vectorization. Using count vectorization, each comment was represented as a vector of term frequencies and the comments were transformed into a matrix of token counts. TF-IDF Vectorization, on the other hand, distinguished between common and unique phrases by assigning weights to terms depending on their significance in the dataset.

Following text vectorization, four machine learning models were selected for sentiment analysis: Multinomial Naive Bayes, Logistic Regression, Random Forest, and Support Vector Machine (SVM). Multinomial Naive Bayes, a probabilistic model based on Bayes' theorem, was applied to model the probability distribution of sentiment labels. Logistic Regression, a linear model used for classification tasks, was employed to predict sentiment labels based on vectorized features.

Random Forest, an ensemble learning method constructing decision trees, captured complex relationships within the dataset to predict sentiment labels. Support Vector Machine (SVM), seeking the hyperplane that best separates data points into different classes, was utilized to classify comments into distinct sentiment categories based on vectorized features.

The models underwent training on a subset of the dataset with labeled sentiment categories, and their performance was evaluated on a separate subset using metrics such as accuracy, precision, recall, and F1 score. To facilitate a comprehensive understanding of the sentiment landscape, the results of sentiment analysis, including predicted and actual labels, were visualized using a heatmap. This visualization provided an intuitive representation of each model's effectiveness in categorizing sentiments expressed in YouTube comments.

In the Sentiment Analysis phase, four distinct machine learning models were applied to categorize YouTube comments into sentiment types: Multinomial Naive Bayes, Logistic Regression, Random Forest, and Support Vector Machine (SVM). Multinomial Naive Bayes, a probabilistic model grounded in Bayes' theorem and well-suited for text classification, was utilized to model the probability distribution of sentiment labels. Logistic Regression, known for its application in both binary and multiclass classification tasks, was employed to predict sentiment labels based on the vectorized features of the comments. The Random Forest model, an ensemble learning method constructing multiple decision trees, captured intricate relationships within the dataset to predict sentiment labels accurately. Additionally, Support Vector Machine (SVM), a model finding the hyperplane best separating data points into different classes, was applied to classify comments into distinct sentiment categories based on vectorized features.

The training and evaluation of these models constituted a critical step in assessing their performance. Each model underwent individual training on a subset of the dataset with labeled sentiment categories. Subsequently, the trained models were evaluated on separate subsets, employing metrics such as accuracy, precision, recall, and F1 score to gauge their effectiveness in categorizing sentiments. The evaluation process provided insights into the strengths and weaknesses of each model in accurately predicting sentiment labels.

Results visualization was carried out through the generation of heatmaps, illustrating the performance of each sentiment analysis model. These heatmaps showcased the predicted labels and actual labels, utilizing color intensity to indicate sentiment scores—red representing negative sentiments and green representing positive sentiments. This visual representation offered an intuitive understanding of how each model categorized sentiments expressed in YouTube comments.

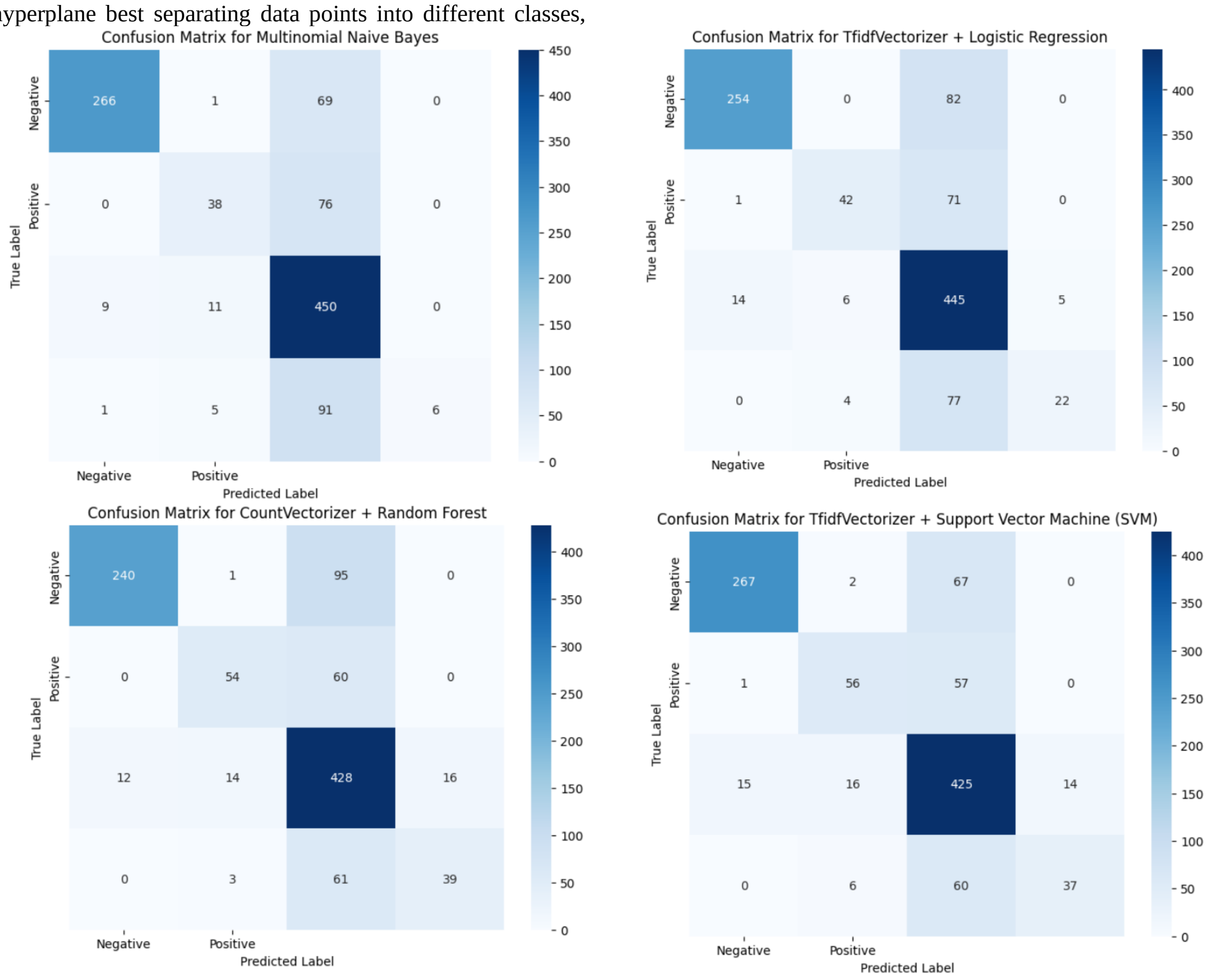

Figure 5 Confusion matrix for different algorithms

The bar chart provides a visual representation of the comparative accuracies achieved by different classifiers in the sentiment analysis of YouTube comments. The four classifiers under consideration are Multinomial Naive Bayes, Tfidf + Logistic Regression, CountVectorizer + Random Forest, and Tfidf + SVM, each contributing to the overall understanding of sentiment categorization. The accuracies for each classifier, denoted by the vertical bars, are as follows: 0.74 for Multinomial Naive Bayes, 0.75 for Tfidf + Logistic Regression, 0.74 for CountVectorizer + Random Forest, and 0.77 for Tfidf + SVM.

The varying heights of the bars illustrate the distinct performance levels of each classifier, with Tfidf + SVM exhibiting the highest accuracy at 0.77. This chart serves as a concise yet informative visual aid, allowing for a quick and direct comparison of the effectiveness of different classifiers in accurately predicting sentiment labels for YouTube comments. The classifiers' performance metrics, measured in terms of accuracy, highlight the nuanced differences in their ability to discern and categorize sentiments expressed in the diverse array of comments within the dataset.

The addition of accuracy values directly on the bars enhances the interpretability of the chart, providing a clear understanding of the quantitative performance metrics associated with each classifier. This comparison aids in the identification of the most effective sentiment analysis approach among the considered classifiers, contributing valuable insights to the evaluation and selection process for sentiment categorization in YouTube comments.

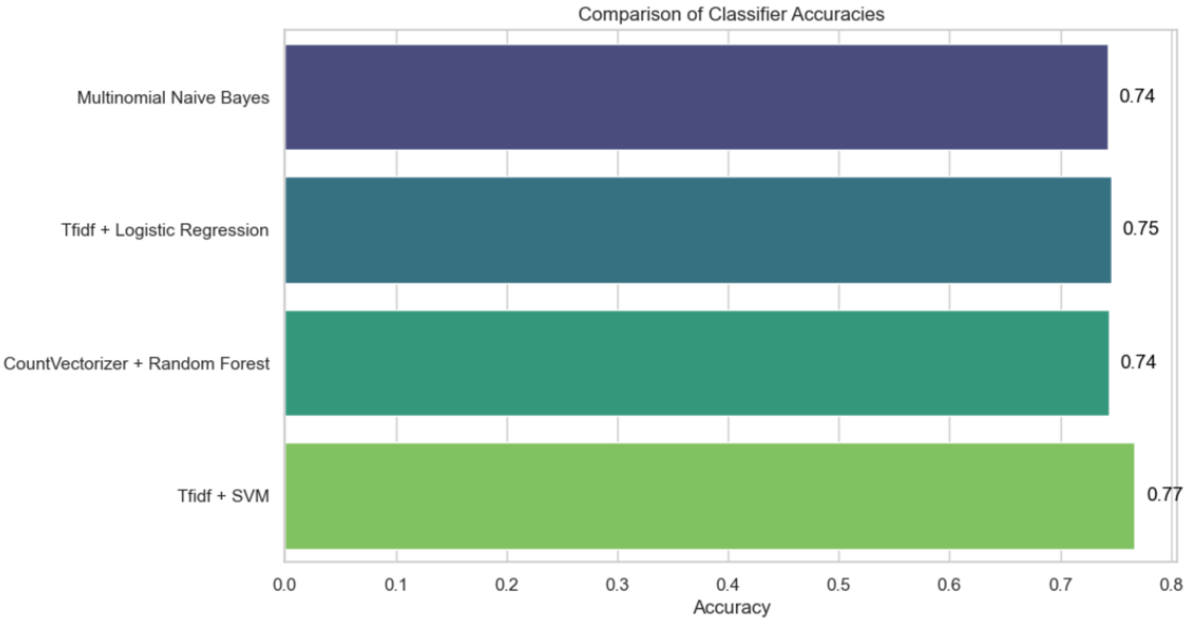

Figure 6 Bar chart for accuracy using different algorithms

The point plot provides an alternative visualization for comparing the accuracies of different classifiers in the sentiment analysis of YouTube comments. In this representation, each classifier is represented by a point along the accuracy scale, offering a more detailed view of the distribution and variability in classifier performance. The data used for this visualization includes the classifiers Multinomial Naive Bayes, Tfidf + Logistic Regression, CountVectorizer + Random Forest, and Tfidf + SVM, along with their corresponding accuracy values: 0.74, 0.75, 0.74, and 0.77, respectively.

The point plot allows for a more granular examination of the accuracy distribution among classifiers, showcasing the subtle differences in their performance. Each point on the plot represents the accuracy of a specific classifier, and the vertical position of the point indicates its corresponding accuracy value. The overall trend and dispersion of points contribute to a nuanced understanding of how each classifier performs in the context of sentiment categorization.

The plot is particularly useful for identifying any patterns, trends, or outliers in the accuracy values, and its horizontal axis, featuring the classifiers, facilitates a direct comparison. The rotation of the x-axis labels ensures better readability, especially when dealing with multiple classifiers. This point plot serves as a complementary visualization to the bar chart, offering an additional perspective on the performance variations among the classifiers, thus enhancing the overall evaluation of sentiment analysis approaches.

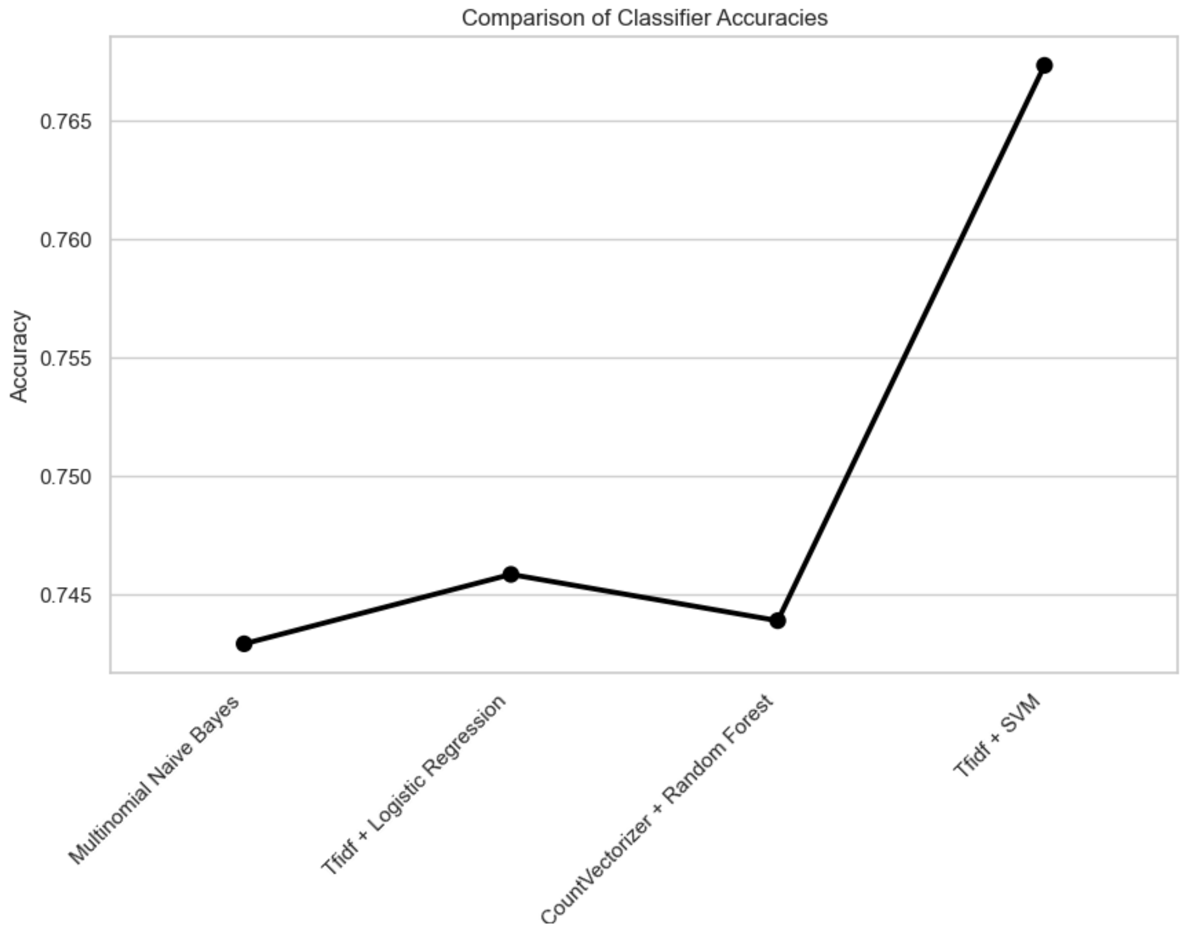

Figure 7 Point chart for accuracy of different classifiers

In the broader context of the research, a Recommendation System was developed based on the sentiment analysis results. This system assigned a score to each video considering the sentiments of its comments, ultimately facilitating the ranking of videos. Higher scores indicated videos with more positive feedback, providing a novel approach to video recommendation grounded in user sentiment.

Algorithm for finding best video

Using the search box, the user initiates the first interaction with the system by conducting a search on a certain topic. A selection of videos pertaining to the query provided are displayed in the search results, which are derived from the YouTube API. These videos, which make up the search result, are then gathered and prepared for additional examination.

The algorithm uses a two-step procedure to select the top movies from the search results. Initially, it extracts every comment related to every video from the search result, and then it stores these comments in an Excel document. The Excel document functions as an all-inclusive database, enumerating the video titles together with the comments that go with them. This collection of remarks serves as the dataset for the sentiment analysis that follows.

A pre-trained machine learning model that was trained on a dataset of YouTube comments tagged with sentiments (positive, negative, or neutral) facilitates the sentiment analysis. The model adds a new column to the Excel sheet that shows the sentiment analysis findings for each remark by predicting its sentiment. This extra layer of data provides an understanding of the general tone that each video's comments have to give.

The method then determines how many comments are overall positive, negative, and neutral for every video. The number of favorable comments divided by the number of negative comments yields the video ratio. This ratio thus becomes an important ranking measure for the videos. Videos having a larger percentage of positive comments are at the top of the list, which is organized in descending order based on their video ratios.

$$video\ ratio = \frac{no.of\ positive\ comments}{no.of\ negative\ comments}$$

A list of videos with their video ratios arranged in order of preference is the ultimate result, which ranks the films in accordance with the opinions stated in the corresponding comments. According to the algorithmic evaluation, the top three videos are the greatest videos since they have the best sentiment ratios. Using machine learning, this method automatically extracts the sentiment dynamics from YouTube comments and uses that information to extract the most well-liked videos on a particular subject from the search results. The complete system offers a streamlined method for finding and suggesting the most interesting and well-received films inside a search context by seamlessly integrating user interaction, sentiment analysis, YouTube API, and machine learning.

4. FUTURE WORK

We can add the feature for the searching filtration. As a student depending on the standard the meaning of searching can be different. For ex. mining of algebra is different for a standard 10th student and a college student. So best video can be different for both the students for same search results.

We'll try to make some changes to this project. We will endeavour to provide users with a positive user experience. As many competitive tests as possible can be added. in order for more pupils to benefit from it. Our system for analysing grades may be made better. For identifying and evaluating weak and strong themes, we can be more precise.

For the learning videos we can extend up to some more standard and can provide more useful material for the same.

5. CONCLUSION

Natural language processing includes the area of sentiment analysis. It focuses on analysing consumer opinions about a certain good or service. Such methods can assist in enhancing the goods, services, or assistance in reversing the course of any subpar goods or services. This research offered a thorough investigation of YouTube video-based sentiment analysis techniques. Easyfy will a complete guide to student for learning something new in a short time. It will help students to find best video on their topics and will help them to boost their learnings. It will be also helpful to the different competitive exam aspirants in their preparation.

6. REFERENCES

[1] Wankhade, M., Rao, A.C.S. & Kulkarni, C. A survey on sentiment analysis methods, applications, and challenges. Artif Intell Rev 55, 5731–5780 (2022). https://doi.org/10.1007/s10462-022-10144-1

[2] C. -C. Lee, Z. Gao and C. -L. Tsai, "BERT-Based Stock Market Sentiment Analysis," 2020 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan), 2020, pp. 1-2, doi: 10.1109/ICCE-Taiwan49838.2020.9258102.

[3] Patil, Ashwini & Gupta, Shiwani. (2021). A Review on Sentiment Analysis Approaches.

[4] Chinnalagu and Durairaj (2021), PeerJ Comput. Sci., DOI 10.7717/peerj-cs.813

[5] C. S. Lee, H. Osop, D. Hoe-Lian Goh, and G. Kelni, "Making sense of comments on YouTube educational videos: A self-directed learning perspective," Online Inf. Rev., vol. 41, no. 5, pp. 611–625, 2017, doi:10.1108/OIR-09-2016-0274.

[6] Sungheetha, Akey, and Rajesh Sharma. "TransCapsule Model for Sentiment Classification." Journal of Artificial Intelligence 2, no. 03 (2020): 163-169.

[7] H. Bhuiyan, J. Ara, R. Bardhan, and M. R. Islam, "Retrieving YouTube video by sentiment analysis on user comment," Proc. 2017 IEEE Int. Conf. Signal Image Process. Appl. ICSIPA 2017, pp. 474–478, 2017, doi: 10.1109/ICSIPA.2017.8120658.

[8] S. Nawaz, M. Rizwan, and M. Rafiq, "Recommendation Of Effectiveness Of Youtube Video Contents By Qualitative Sentiment Analysis Of Its Comments And Replies," Pak. J. Sci., vol. 71(4 Suppl,pp. 91–97, 2019.

[9] E. Mulholland, P. M. Kevitt, and T. Lunney, "Analysing Emotional Sentiment in People ' s YouTube," vol. 2, pp. 181–188, 2017, doi:10.1007/978-3-319-55834-9.

[10] M. Park, M. Naaman, and J. Berger, "A data-driven study of view duration on YouTube," Proc. 10th Int. Conf. Web Soc. Media, ICWSM 2016, no. Icwsm, pp. 651–654, 2016.

[11] J. Savigny and A. Purwarianti, "Emotion classification on youtube comments using word embedding," Proc. - 2017 Int. Conf. Adv. Informatics Concepts, Theory Appl. ICAICTA 2017, pp. 1–5, 2017, doi: 10.1109/ICAICTA.2017.8090986.

[12] I. Perikos and I. Hatzilygeroudis, "Aspect based sentiment analysis in social media with classifier ensembles," Proc. - 16th IEEE/ACIS Int. Conf. Comput. Inf. Sci. ICIS 2017, pp. 273–278, 2017, doi: 10.1109/ICIS.2017.7960005.

[13] R. Benkhelifa, "Opinion Extraction and Classification of Real-Time YouTube Cooking Recipes Comments," pp. 395–404, 2018, doi: 10.1007/978-3-319-74690-6

[14] H. Jelodar et al., "A NLP framework based on meaningful latent-topic detection and sentiment analysis via fuzzy lattice reasoning on youtube comments," pp. 4155–4181, 2021.

[15] S. Paliwal, S. Kumar Khatri and M. Sharma, "Sentiment Analysis and Prediction Using Neural Networks," 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), Coimbatore, 2018, pp. 1035-1042, doi: 10.1109/ICIRCA.2018.8597358.

[16] M. Aufar, R. Andreswari and D. Pramesti, "Sentiment Analysis on Youtube Social Media Using Decision Tree and Random Forest Algorithm: A Case Study," 2020 International Conference on Data Science and Its Applications (ICoDSA), Bandung, Indonesia, 2020, pp. 1-7, doi: 10.1109/ICoDSA50139.2020.9213078

[17] H. AlSalman, "An Improved Approach for Sentiment Analysis of Arabic Tweets in Twitter Social Media," 2020 3rd International Conference on Computer Applications & Information Security (ICCAIS), Riyadh, Saudi Arabia, 2020, pp. 1-4, doi: 10.1109/ICCAIS48893.2020.9096850

[18] A. Mitra, "Sentiment Analysis Using Machine Learning Approaches (Lexicon based on movie review dataset)," J. Ubiquitous Comput. Commun. Technol., vol. 2, no. 3, pp. 145–152, 2020, doi: 10.36548/jucct.2020.3.004.

[19] . Yang, Y. Li, J. Wang and R. S. Sherratt, "Sentiment Analysis for ECommerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning," in IEEE Access, vol. 8, pp. 23522-23530, 2020, doi: 10.1109/ACCESS.2020.2969854.

[20] Chauhan, G. S., & Meena, Y. K. (2019). YouTube video ranking by aspect-based sentiment analysis on user feedback. In Soft Computing and Signal Processing: Proceedings of ICSCSP 2018, Volume 1 (pp. 63-71). Springer Singapore.