

Enhancing Viewer Experience on YouTube: A Machine Learning Approach for Assessing Tutorial Video Quality

Jai Rizvi, Tony Gao & Cooper Yang

Received September 17, 2023

Accepted March 21, 2024

Electronic access March 31, 2024

Finding high-quality educational content on YouTube can be a time-consuming and frustrating task for viewers seeking tutorials on various subjects. In this research paper, we present a machine learning model designed to assess the quality of YouTube tutorials by analyzing metrics such as likes, dislikes, comments, and views, and then assigning a rating out of five stars. By leveraging this model, viewers can streamline their search process and enhance their overall experience by easily identifying the most valuable and informative tutorials. We trained our model on a large dataset of 50 different YouTube tutorials, and when we matched its predictions to Chat GPT's ratings out of 5 based on the YouTube video transcripts, we found that it was quite accurate on the testing data, having a mean squared error (MSE) of 0.06. We asked Chat GPT to rate the YouTube tutorial transcript on a scale from 1 to 5 based on efficiency, quality, and use of demonstrations. This demonstrates the model's effectiveness, consistently matching Chat GPT's human-like ratings and thereby reinforcing its reliability in determining the quality of tutorials. Chat GPT's reviews are trained on actual reviews given by people online, therefore it's safe to assume that Chat GPT can very closely represent an average human's opinion when provided the script for the YouTube video. Additionally, our research unveils an unexpected finding—the model's versatility extends its applicability beyond tutorials to regular entertainment videos, revolutionizing video content assessment across diverse genres and types on YouTube, further improving the viewer experience.

Introduction

YouTube has become a vast repository of educational and entertainment content, with tutorials on various subjects being a significant part of its offerings. However, the sheer volume of videos presents a challenge for viewers, as not all tutorials provide the same level of value, information, and quality. Consequently, finding high-quality tutorials can be a daunting and time-consuming process. This study aims to address this issue by introducing a machine learning model capable of objectively assessing the quality of YouTube tutorials and providing the video rating out of five stars. Furthermore, the model's secondary goal is to assist creators with optimizing their videos. The rating may give creators a better framework to create videos, and also help them easily compare different videos and how they meet consumer satisfaction.

Prior research has explored analyzing comments specifically on YouTube, but few studies have focused specifically on evaluating the quality of individual tutorials based on all of its metrics. Siersdorfer et., al, (2010)¹ proposed a model somewhat similar to ours in the fact that it analyzed the positivity of the comments on a YouTube video, and then used that to figure out the overall rating of the video. The study analyzed a vast dataset of over 6 million comments on 67,000 YouTube videos. The researchers investigated the relationships between comments, views, comment ratings, and topic categories. Additionally, they

examined how the sentiment expressed in comments influenced their ratings, using the SentiWordNet thesaurus.

R.Singh & A. Tiwari, (2021)² scraped YouTube comments and measured the attitude of the user towards the video they commented on. This study addresses the challenge of extracting meaningful trends from YouTube comments due to their low information consistency and quality. The researchers perform sentiment analysis on comments related to popular topics using various machine learning algorithms. They aim to identify trends, seasonality, and forecasts in sentiments to understand the influence of real-world events on public sentiments.

Ashgar, et., al. (2023)³ used another sentiment analysis model to analyze the opinions about certain YouTube videos based on the comments left by other users. They provide a brief summary in order to demonstrate the techniques used to run a sentiment analysis model on youtube comments. One method in particular categorized YouTube comments into 4 different types: short syllable comments, advertisements, negative criticism, and absurd ranting. This study is outdated, so we found it of use to update their findings in the new online landscape.

The relevance and significance of this research lie in its potential to significantly enhance the viewer experience on YouTube. By simplifying the process of finding high-quality tutorials, our model allows users to make more informed decisions about the content they consume. This, in turn, benefits content creators by directing more viewers towards valuable tutorials. Moreover,

our model's unexpected versatility, capable of assessing regular videos as well, opens up new possibilities for content evaluation on the platform.

In the subsequent sections, we will delve into the gathering data for and training of our machine learning model, the variables used in the training datasets, and the evaluation of the model's performance. We will present testing on data points that are beyond our training and testing data, which will demonstrate the model's accuracy and reliability in assigning ratings to tutorials. Furthermore, we will discuss the broader implications of our model's potential applications beyond tutorials, paving the way for more efficient content discovery on YouTube. Overall, this research contributes a valuable tool to the YouTube community, promoting a more rewarding and efficient viewing experience for users and creators alike.

Results

Now that we collected all of the data to train our model, we employed various machine learning models to perform sentiment analysis on YouTube comments. The primary model utilized was the Random Forest Regression model, which yielded an average mean squared error (MSE) of 0.06, indicating its relatively high predictive performance. Our testing data was set to 25% of our overall data. We used MSE because when the model predicted the star rating for each video, came out to be a decimal, something like 3.24. To combat this, we used the MSE to test the accuracy, so it could measure how far the prediction was from the actual star rating. Following closely was the Linear Regression model, generating a MSE of 0.195. Other models were also considered which didn't have the best accuracy, including Logistic Regression with a 60% accuracy rate, SVM with 20% accuracy, Decision Tree with 10% accuracy, Naïve Bayes with 20% accuracy, and Neural Network with a 40% accuracy. These results offer insights into the comparative effectiveness of different models in capturing the sentiment patterns within the YouTube comments dataset.

In Table 1, we present the standard deviation values observed across each star rating category in the Random Forest Regression Model. The standard deviation is a measure of the dispersion or variability of the data points within each category. For example, if we look at figure 4 we can see that the standard deviation across 1-star ratings is 0.14572, which is the absolute value of difference between the predicted star rating and the Chat GPT star rating. This means that on average, when we compare the 1-star rating of Chat GPT to the predicted star rating of our model, it is on average 0.14572 off. Our Random Forest Regression model's total standard deviation was 0.236, which means that, on average, the predicted star ratings for comments in the YouTube dataset differed from the actual star ratings given by Chat GPT by 0.236. This indicates that the model's predictions were, on average, approximately 0.236 stars away from the true ratings

assigned by Chat GPT. The total standard deviation value gives us an overall measure of the model's accuracy in predicting star ratings for comments on the platform. Lower values for the standard deviation indicate more accurate predictions, while higher values suggest more variability or less precision in the model's estimations.

Table 1: The standard deviation of each star rating

Standard deviation across star ratings				
1 star	2 stars	3 stars	4 stars	5 stars
0.14572	0.17302	0.21043	0.21360	0.13203

When we got a summary on our linear regression model, we found that the R-squared was 0.945. The R-squared value of 0.945 obtained from the Linear Regression model indicates an exceptionally strong correlation between the features (independent variables) and the target variable (star rating out of 5).

We now present results from the linear regression summary, delving into each coefficient and p-value for each independent variable, providing insights into the significance and influence of these variables in the star ratings out of 5 provided by Chat GPT.

1. View to like ratio:
Coefficient: 0.0086
P-value: 0.008

The coefficient of the view to like ratio is 0.0086, indicating that for every one-unit increase in the view to like ratio, the star rating out of 5 is estimated to increase by 0.0086 units, assuming all other variables remain constant. The p-value of 0.008 is less than the significance level of 0.05, indicating that this coefficient is quite significant. Therefore, the view to like ratio has a meaningful impact on the star rating.

2. View to comment ratio:
Coefficient: 1.807e-05
P-value: 0.910

The coefficient of the view to comment ratio is 1.807e-05, indicating that for every one-unit increase in this variable, the star rating out of 5 is estimated to increase by 1.807e-05 units, which is a very small value. The p-value of 0.910 is greater than the significance level, implying that this coefficient is not statistically significant. Consequently, the view to comment ratio may not have a significant influence on the star rating out of 5, and its inclusion in the model may not be all that critical.

3. View to dislikes ratio:
Coefficient: 2.531e-05
P-value: 0.861

The coefficient of the view to dislikes ratio is 2.531e-05, indicating that for every one-unit increase in the view to dislikes ratio, the star rating out of 5 is estimated to increase by 1.807e-05 units, which is again a very small value, similar to the previous variable. Similarly, the p-value of 0.861 is greater than the significance level, indicating that this coefficient is not very significant.

4. Like to dislike ratio:
Coefficient: 0.0014
P-value: 0.734

The coefficient of the likes to dislikes ratio is 0.0014, meaning that for every one-unit increase in the likes-to-dislikes ratio, the star rating out of 5 is estimated to increase by 0.0014 units. The p-value of 0.734 is greater than the significance level, indicating that this coefficient is not statistically significant. As a result, the likes-to-dislikes ratio may not be a strong predictor of sentiment in YouTube comments, and its inclusion in the model may not significantly contribute to the star rating out of 5.

5. Comment’s mean sentiment score:
Coefficient: 9.8272
P-value: 0.000

The coefficient of the comment’s mean sentiment score is 9.8272, meaning that for every one-unit increase in the comment’s mean sentiment score ratio, the star rating out of 5 is estimated to increase by 9.8272 units. The p-value of 0.000 is far below the significance level, indicating that this coefficient is highly statistically significant. Hence, the comment’s mean sentiment score has a substantial impact on the star rating out of 5.

In conclusion, the linear regression summary provides valuable insights into the significance of each independent variable in predicting the star rating out of 5. The results indicate that the view to like ratio, and particularly the comment’s mean value, are highly significant predictors of the star rating. On the other hand, the view-to-comment ratio and view-to-dislikes ratio are not the most significant predictors, while the likes to dislikes ratio has only a marginal impact. These findings can aid content creators, marketers, and platform administrators in understanding the factors that influence how the public feels about a certain YouTube video. Figure 1 presents a bar chart depicting the individual importance of each independent variable in the analysis, represented by their respective coefficients. Unlike the linear regression model’s summary, which considers the joint impact of all variables while controlling for other factors, this chart highlights the isolated influence of each variable without the “noise” introduced by other variables’ interactions. By isolating

the variables’ impacts, the bar chart allows us to gain a clearer understanding of the relative importance of each independent variable in influencing the dependent variable, which in this case is the star rating provided by Chat GPT. It is worth noting that we dropped an independent variable called engagement rate, which was calculated by doing “likes + comments / duration of the video”, which turned out to be not that helpful, and having very little impact on the dependent variable, which used to be reflected in this bar chart.

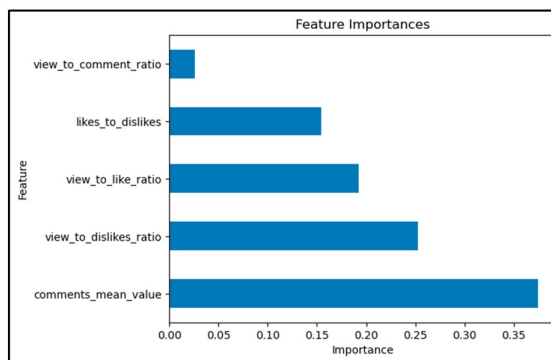


Fig. 1 A bar chart that represents the amount of importance for each independent variable on the dependent variable

In Figure 2 we have a heatmap, which represents the correlation matrix between the star rating, the dependant and the independent variables (view_to_like_ratio, view_to_comment_ratio, view_to_dislikes_ratio, likes_to_dislikes, and comments_mean_value). The correlation values range from -1 to 1, where -1 indicates a perfect negative correlation, 1 indicates a perfect positive correlation, and 0 indicates no correlation between the variables. Analyzing the heatmap, we will now observe the following relationships to the star rating:

- Strong Negative Correlation with view_to_like_ratio to star rating (-0.526249):
- Moderate Positive Correlation with view_to_comment_ratio to star rating (0.236464):
- Moderate Positive Correlation with view_to_dislikes_ratio to star rating (0.472207):
- Moderate Positive Correlation with likes_to_dislikes to star rating (0.470469):
- Strong Positive Correlation with comments_mean_value to star rating (0.850395):

The star rating shows a strong positive correlation with the mean sentiment score of comments. As the average sentiment expressed in comments becomes more positive, the star rating tends to increase significantly. This indicates that comments with more positive sentiments are associated with higher star ratings for the video.

In conclusion, the view-to-like ratio, view-to-comment ratio,

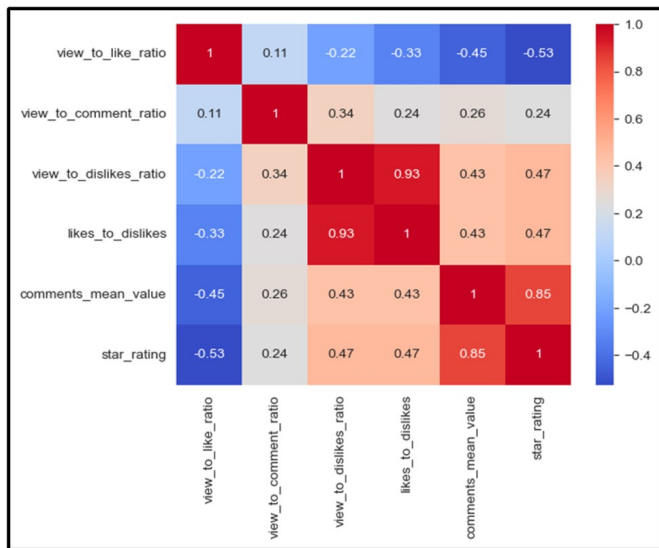


Fig. 2 A heatmap that represents correlations across the variables

view-to-dislikes ratio, likes-to-dislikes ratio, and the mean sentiment score of comments are all significantly related to the star rating. The insights gained from this analysis can be utilized to understand the factors influencing user sentiments and engagement on YouTube, helping content creators and platform administrators to optimize content and foster a more positive user experience on the platform.

In Figure 3, we observe the distribution of star ratings ranging from 1 to 5, as generated by Chat GPT. An important observation from the distribution is that the ratings of 3 and 4 were substantially higher in frequency compared to ratings 1, 2, and 5. This uneven distribution of ratings implies that a considerable proportion of Chat GPT's ratings provided feedback with ratings of 3 and 4, while ratings 1, 2, and 5 were less prevalent in the dataset.

The uneven distribution of ratings may have implications for the model's predictions and potential biases in its performance. Since the data is weighted towards ratings 3 and 4, the model may be more inclined to predict outcomes closer to these ratings. As a result, the model's predictions could tend to be biased towards the moderate and positive sentiment levels represented by ratings 3 and 4, respectively. For example, when asked to predict the sentiment of a new YouTube comment, the model may be more likely to assign a rating close to 3 or 4 due to the higher frequency of these ratings in the training data. This might result in an underestimation of negative sentiments (ratings 1 and 2) and an overestimation of positive sentiments (ratings 4 and 5). We did not address this issue, but a way we could have worked around this imbalance in data is giving more data for other data points so that the data will be even.

It is important to be aware of this potential bias when in-

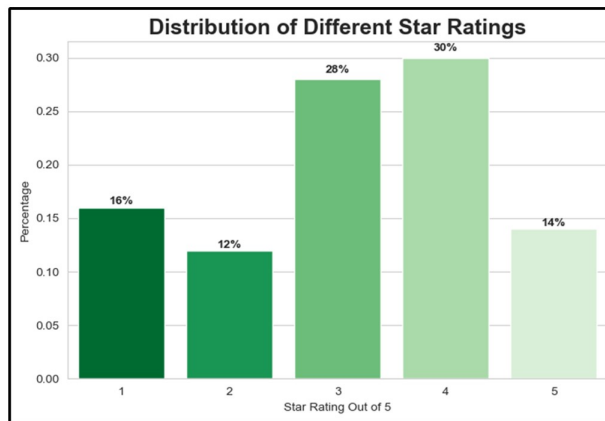


Fig. 3 A bar graph displaying the frequency of different star ratings

terpreting the model's predictions. While the model may be accurate in predicting sentiments close to ratings 3 and 4, its performance on extreme ratings (1 and 5) could be less reliable due to the limited representation of such ratings in the training data.

In Table 2, we present the results of testing our Random Forest Regression model on a new dataset, which consists of four different videos with their corresponding independent variables, Chat GPT's star ratings, and predicted star ratings by our model. The primary objective of this analysis was to evaluate the model's performance on data that it had not encountered during training, enabling us to assess its ability to generalize and make accurate predictions on unseen samples. After evaluating various machine learning models, we found that the Random Forest Regression model yielded the best results with a mean squared error (MSE) of 0.06. Linear regression provided the second-best performance, achieving an MSE of 0.195. Other models were also assessed, including logistic regression with 60% accuracy, support vector machines (SVM) with 20% accuracy, decision trees with 10% accuracy, Naïve Bayes with 20% accuracy, and neural networks with 40% accuracy. The Random Forest Regression model outperformed all other models in terms of accuracy and minimizing the mean squared error.

Upon analyzing the results, we observe that the predicted star ratings are quite close to Chat GPT's star ratings, indicating that the model is capable of making reasonably accurate predictions on unseen data. For example, in the first video represented by video link "NRnaMCNOK7Y," the actual star rating is 4, and the model's prediction is 4.32, which is a close approximation. Similarly, for video "REas9cmjlic," the actual star rating is 3, and the model predicts a rating of 3.22. The model's average deviation from Chat GPT's star rating is 0.22, which is not bad at all. We made sure that the model was not overfitting because our testing data held no similarity to the training data, so it wasn't possible that the model could generalize our testing data

Table 2: A bar graph displaying the frequency of different star ratings

video_link	view_to_like_ratio	view_to_comment	view_to_dislikes_ratio	likes_to_dislikes	comments_mean_value	star_rating	predicted star rating
NRnaMCNOK7Y	38.73420261	643.6333333	1838.952381	47.47619048	0.4387	4	4.32
REas9cmjlic	31.61259375	480.5714964	3010.723214	95.23809524	0.1374	3	3.22
d_XgMTDkJP8	61.57852883	61.21343874	192.3850932	3.757763975	-0.0827	1	1.04
q4EGeonO6YI	14.22762162	463.6036988	1159.519824	81.49779736	0.3822	4	4.3

into our training data.

Overall, the performance of the Random Forest Regression model on the new dataset appears to be quite promising, as it successfully provides close approximations of the actual star ratings for most videos. When we round each prediction to the nearest whole number, all of them end up being correct.

Discussion

The significance of the results obtained in this research paper lies in the development of a machine learning model capable of assessing the quality of YouTube tutorials and providing a star rating out of five. For the average consumer, this rating can be an effective tool when determining if a tutorial is worthwhile. The training set's scoring system derived from Chat GPT provides a blanket scoring system that originates from online user media, meaning it covers a spectrum of metrics that one would use when critiquing a tutorial. By analyzing metrics such as likes, dislikes, comments, views, and sentiment scores, the model can efficiently evaluate tutorial videos, streamlining the search process for viewers seeking high-quality educational content. This model combines both subjective opinions and objective statistics, allowing it to derive an "average" opinion for any video, which will allow the user to gauge which videos to watch. The high performance of the model, with a mean squared error of 0.06, demonstrates its accuracy and reliability in predicting ratings, matching closely with Chat GPT's human-like ratings. This tool offers a valuable resource to YouTube users, helping them identify the most valuable and informative tutorials, thereby enhancing their viewing experience. Furthermore, the research shows the model's versatility, extending its applicability beyond tutorials to regular entertainment videos. The insights gained from the analysis of the model's performance, the significance of individual variables, and the correlation between features and ratings can assist content creators, marketers, and platform administrators in understanding user sentiments, optimizing content, and fostering a more positive user experience on the platform. Overall, the results contribute a valuable tool to the YouTube community, promoting more informed decisions for viewers and creators alike, ultimately enhancing the overall viewer experience on the platform.

Next steps:

- **Incorporate more variables:** Adding additional relevant variables, such as video duration, likes on each comment to gauge agreement, or even the video's engagement rate, can provide more comprehensive insights into the video's quality. These variables can help capture more nuanced aspects of user satisfaction and contribute to a more accurate assessment.
- **Increase dataset size:** Expanding the dataset beyond the initial 50 tutorials would be beneficial for enhancing the model's accuracy. A larger dataset can provide a more diverse and representative sample, enabling the model to generalize better to a wider range of tutorials and video types.
- **Optimize comment sentiment analysis keyword search:** a broader variety of both positive and negative vocabulary incorporated into the keyword search will allow the tool to gather a larger count of comments with positive or negative reviews; comments with sentiment will boost accuracy.
- **Extend to all YouTube videos:** As suggested, broadening the model's scope to encompass all types of YouTube videos, not just tutorials, can revolutionize content assessment on the platform. This expansion could include various genres, such as entertainment, vlogs, reviews, and more, offering users a comprehensive tool to discover high-quality content across diverse categories. The model just needs to have a wider range of data for this to be achieved, and the same principles used for rating the quality of the tutorials can be applied to diverse genres.
- **Real-world implementation:** Building a Chrome extension that automatically scrapes and rates YouTube videos on a user's home page would be a practical and effective way to improve their viewing experience. This extension could display star ratings next to each video, making it easier for users to identify valuable content and make informed choices without the need to review multiple videos. The YouTube API facilitates the scraping of video data, and all creators on the platform consequently allow it when they produce content on the site. Besides dislikes, these metrics are publicly available under the video, which eliminates concerns about data privacy. Dislikes are a more morally

ambiguous area, although other chrome extensions that only extrapolate dislikes have not encountered such issues involving platform policies or user content, so it's safe to do so.

- Human video evaluation: Moving away from relying solely on Chat GPT's transcript-based ratings, human evaluation of each individual video can provide a more accurate and comprehensive assessment. By watching the videos, evaluators can consider various elements such as clarity of explanations, audio and visual quality, engagement with the audience, practical demonstrations, and overall production value.

This model builds the foundation to expand into a valuable tool, as it has been shown that scoring is correlated with these statistics. The dataset is a venture into this area, and shows the potential of the model when it meets a larger dataset. The choice of using tutorials to center the model on was to take advantage of existing comment sentiment analysis models as tutorial comments are typically either positive or negative and not as reliant on the content itself; advancing comment sentiment analysis would be the next step in being more precise with other genres of content.

Methodology

To train the model we used 5 features to predict the target, which was the rating out of 5 stars from Chat-GPT. Its ratings are derived from actual reviews given by people online, therefore it's safe to use it as a benchmark for the average satisfaction that it would provide a person. The features consisted of the view to like ratio, the view to number of comments ratio, the view to dislikes ratio, the likes to dislikes ratio and the comment mean sentiment score. The selection of these metrics were based on two reasons: they are easy to access and they have a high R^2 value, meaning they have high correlation to the final scoring. When a user finds a video that benefits them, be it entertaining or helpful, they will show their satisfaction by interacting with the video, such as liking or commenting. Choosing it to be a ratio with views allows the metrics to be scalable; it accounts for a video being good yet obscure.

To scrape the likes, dislikes and number of comments from the YouTube videos, we had to use the YouTube API. The YouTube API scraped the metrics from the video given. We wanted to get these metrics because we thought that seeing how many people viewed the video and finding out how many of those people liked, disliked, or commented on the video was valuable insight on giving the video a rating out of 5 stars. All we had to do to get these metrics from each YouTube video was to assign a string variable to the link of the video we want to use, and then run a scraping function we wrote that would get the likes,

dislikes, views and number of comments. Then once we got those metrics the function would turn all of the variables into the ratios we see as the features. The reason why we used the ratios was so that all of the videos would be proportional to each other. The code then took all of the ratios and made a new row into an existing csv we had, and we had to do this for each video. Figure 4 shows the csv file we used. We just had to put the link of the video in, and then it would add a new row. It is worth noting that we had to collect the dislikes manually because YouTube removed dislikes from every video, which made it unavailable to scrape from the YouTube API. We used a chrome extension called 'Return YouTube Dislikes' so that it would reveal the hidden dislike count and we could manually assign it to a variable in our function. Dislike statistics appear on every video and are merely hidden, therefore the collection of the dislike count does not introduce further bias.

To scrape the actual comments and get the mean sentiment score across the whole video's comment section we again used the YouTube API. The reason why we wanted to get the comments from a YouTube video was because we wanted to gauge how the audience felt about the video. To retrieve all the comments in a given list of videos, we used the `youtube.commentThreads()` function to scrape it directly off of the platform. To retrieve just the first 50 comments, we used a `max_comments` variable and a for loop to stop retrieving comments once we reach the desired amount. Once we collected the comments we compiled them into a csv file. To then analyze the comments, we used a natural language processing (NLP) model called Vader Lexicon to analyze each comment and give every comment a score from -1 to 1. The most negative a comment could be is -1 and the most positive a comment could be is 1. Figure 5 illustrates what the Vader Lexicon model produced. These are just the first few comments out of 50 comments, and we can see in the 'compound' column the model gives each comment a score from -1 to 1. The bottom 4 video's comments are labeled as positive, while the first comment is a neutral comment. What we did was take all of the comments' compound scores and get the mean of all of them, and then that would be the comments' mean sentiment score for the video. Figure 6 displays 4 videos' comments mean sentiment scores. We can see from this table that videos 1 and 4 have a majority of positive comments, and the mean compound sentiment score reflects that. Video 2, however, has a lower mean sentiment score, nearing toward more of a neutral group of comments, while video 3 is in the negatives, which means that there were many negative comments, more than the amount of positive comments. Getting a sense of how the audience feels about a video is crucial to finding out how good a tutorial is because the comments on a YouTube tutorial should reflect what the audience took away and how they felt the tutorial was.

To get the star rating out of 5 for each video (the target), we initially thought that we should watch every video and give it

	video_link	view_to_like_ratio	view_to_comment_ratio	view_to_dislikes_ratio	likes_to_dislikes
0	https://www.youtube.com/watch?v=94UHCEmprCY	53.477432	1763.787330	2498.698718	46.724359
1	https://www.youtube.com/watch?v=7Ron6MN45LY	84.100584	925.571433	938.741148	11.162124
2	https://www.youtube.com/watch?v=ZHDai11Xga0	18.183626	450.637681	4442.000000	244.285714
3	https://www.youtube.com/watch?v=PXaLc9AYlcg	40.241665	976.668121	3661.422672	90.985865
4	https://www.youtube.com/watch?v=UUaNiEmwEh0	30.344953	450.969880	2023.270270	66.675676
5	https://www.youtube.com/watch?v=hbxQw4LQwws	44.001992	1463.993976	4585.339623	104.207547

Fig. 4 The CSV file that was used to train our data

	Comment	Positive	Negative	Neutral	Compound
0	for x in range(1, 10): if x%2==0: ...	0.000	0.000	1.000	0.0000
1	thanks for your clear explanation	0.647	0.000	0.353	0.6705
2	I was doing so great up to this point, I don't...	0.344	0.000	0.656	0.7384
3	I've looked everywhere online and I can't find...	0.138	0.000	0.862	0.7998
4	Love the example you showed with the email. Ma...	0.389	0.000	0.611	0.7906

Fig. 5 The results of the Vader Lexicon model

comments_mean_value
0.4387
0.1374
-0.0827
0.3822

Fig. 6 The average sentiment scores across comments in 4 videos

References

- 1 S. Siersdorfer, Proceedings of the 19th international conference on World wide web.
- 2 R. Singh1 and A. Tiwari, *International Journal of Scientific Research in Engineering and Management (IJSREM)*.
- 3 M. Asghar, S. Ahmad, A. Marwat and F. Kundi, *Sentiment Analysis on YouTube: A Brief Survey*, <https://arxiv.org/ftp/arxiv/papers/1511/1511.09142.pdf>, Retrieved July 22, 2023, from.

a star rating out of 5. We soon found out that this was very inefficient and time consuming, because watching every single video on our dataset would take days. There would also be a heavy convenience bias involved because there would be various opportunities to influence the rating based off of the features, like view to like ratio or like to dislike ratio. We ended up choosing Chat GPT to rate the YouTube video. This way there would be no bias when rating the video, as Chat GPT is not capable of having a bias. We gave the YouTube video's transcript to Chat GPT, and then asked it to give the video a star rating out of 5 based off of the video's efficiency, quality, and use of demonstrations. We also asked Chat GPT to keep the star ratings as whole numbers, so that way it would be easier for our model to predict, and also easier for the user to interpret.