09

Effect of Comment Sentiments on the Viewership Performance of Video Channels: A YouTube Sentiment Analysis Study

[1]Sanjay Goswami [0000-0002-4141-4934] and Rahul Bhowmik

Institute of Computing and Analytics, NSHM Knowledge Campus, Kolkata, India
[1]sanjay.goswami@nshm.com

Abstract

The authors investigated the relationship between the sentiments expressed in YouTube comments and the viewership performance of the videos, channels, and topics. The study explores the connection between YouTube comment sentiments and viewership performance at multiple levels (videos, channels, topics). The authors used a 2023 dataset to analyze this relationship and conducted a comparative study of machine learning (ML) models for sentiment classification. They concluded that while comment sentiments impact performance metrics, they are not definitive indicators of performance due to the multifaceted nature of success on YouTube. Additionally, large language models (LLMs) demonstrated superior performance in sentiment classification compared to other ML models. Future research could involve refining sentiment metrics or examining other YouTube performance indicators.

Keywords: YouTube sentiment analysis, viewership performance, machine learning models, large language models (LLMs), social media engagement.

1. Introduction

With the rapid increase in user-generated content on YouTube, analyzing the emotions expressed in comments is crucial for content creators, marketers, and platform managers. This study explores how the sentiments in YouTube comments relate to the videos, channels, and topics they are associated with. Sentiment analysis has been a widely researched topic, particularly in online platforms, where understanding user

opinions helps drive engagement and decision-making (Athar 2014; Saif et al. 2012). The research takes a two-pronged approach: it starts with an *exploratory analysis* to identify patterns in sentiment across different types of content, followed by *predictive modeling* to evaluate various machine learning techniques for analyzing comment sentiment. These models are compared based on their performance, helping to determine the most reliable methods for sentiment analysis in this setting. Previous studies have shown that advanced models, including dynamic linguistic approaches, effectively capture sentiments in context (Poria et al. 2015; Sobhani et al. 2016). The findings highlight the complex relationship between viewer sentiment and YouTube content, providing valuable insights that can guide creators and managers in improving engagement and user satisfaction. By focusing on sentiment analysis, the study contributes to a deeper understanding of user feedback on online platforms, emphasizing its importance for creating better user experiences and making more informed decisions.

2. Literature Review

Researchers have widely explored sentiment analysis on social networks such as Twitter and YouTube by examining comments, tweets, and other user-generated content to better understand public interactions on these platforms. For instance, Siersdorfer et al. (2010) conducted an extensive analysis of over six million comments from 67,000 YouTube videos, examining the relationships between comments, view counts, comment ratings, and topic categories. Their work led to the development of prediction models that accurately forecast the ratings of unrated comments based on previously rated ones.

Similarly, Pang et al. (2002) studied 2,053 movie reviews from IMDb to evaluate whether sentiment analysis could be treated as topic-based text classification. Their findings revealed that machine learning approaches, such as Naive Bayes (NB) and Support Vector Machines (SVMs), significantly outperformed manual classification methods, highlighting the effectiveness of automated techniques in sentiment analysis. The accuracy of sentiment classification often falls short compared to traditional topic-based text categorization when employing machine learning techniques. This disparity arises from the inherent complexity of sentiment analysis, as reviews frequently contain a mix of positive and negative expressions, making it challenging for models to accurately determine the overall emotional tone (Pang & Lee, 2008). Addressing this challenge, Shree & Brolin (2019) focused on analyzing YouTube comments and proposed an unsupervised lexicon-based approach to detect sentiment polarity. Their methodology involved creating a specialized social media lexicon to capture user sentiments and opinions effectively. Despite its innovative design, the study revealed a significant limitation: the recall of negative sentiments was notably lower than that of positive sentiments. This discrepancy was attributed to the wide variety of linguistic expressions used to articulate frustration and dissatisfaction, which are harder to standardize and interpret in sentiment analysis models.

Research has also focused on sentiment analysis within social networks like Twitter, uncovering links between individuals, moods, and events across social, political, cultural, and economic domains (Kramer et al., 2014). Building on this, Kowcika

et al. (2019) proposed a system to analyze sentiments in tweets specifically related to the smartphone market. Their approach integrates an efficient scoring mechanism to predict users' ages and employs a well-trained Naïve Bayes (NB) Classifier to determine users' genders. Furthermore, the system incorporates a Sentiment Classifier Model to assign sentiment labels to tweets, providing a comprehensive framework for understanding public opinions in this context.

Balog et al. (2013) developed a system designed to collect and analyze sentiments in tweets related to the smartphone market. Their approach included an efficient rating mechanism to predict users' ages, offering a novel perspective on demographic analysis in sentiment research.

Kouloumpis et al. (2011) extended the exploration of sentiment analysis by investigating the effectiveness of linguistic features in detecting sentiment within Twitter messages. Their study assessed existing lexical resources and features that address the informal and creative language prevalent in microblogging platforms, emphasizing the need for tailored methods in social media sentiment analysis. Additionally, Mishne (2006) conducted sentiment analysis on web text by examining blog posts, highlighting the potential of blogs as rich sources of public opinion and sentiment data.

Riboni's (2004) significant contribution to website classification is detailed in his study, Feature Selection for Website Classification. This research focused on an experimental dataset comprising 8,000 documents sourced from 10 Yahoo! categories. The study employed Kernel Perception and Naive Bayes classifiers to evaluate classification performance. A key finding was the effectiveness of dimensionality reduction techniques, which enhanced classification accuracy. Riboni also introduced an innovative structure-oriented weighing technique that further improved the process. Additionally, the research proposed a novel approach for representing linked pages by leveraging local information, thereby enabling hypertext categorization suitable for real-time applications.

Similarly, Frank et al. (2003) proposed a correction method by adjusting attribute priors in their classification work. This correction, implemented as an additional data normalization step, significantly improved the area under the ROC curve. They demonstrated that the modified version of Multinomial Naive Bayes (MNB) is closely related to the simple centroid-based classifier and conducted empirical comparisons of the two methods.

Diana Maynard et al. (2014) explored sentiment analysis of social media using a multimodal approach in their paper. They focused on assisting archivists in selecting material for inclusion in a social media archive aimed at preserving community memories, moving towards structured preservation around semantic categories. Their rule-based textual approach addressed challenges specific to social media, such as noisy and ungrammatical text, the use of swear words, and sarcasm. Additionally, Athar's 2014 work on sentiment analysis of scientific citations is mentioned, indicating its relevance in the broader field of sentiment analysis research (Athar, 2014).

3. Methodology

The study was conducted through a series of well-defined steps. First, a central repository of channels was constructed to serve as the foundational dataset. Next,

data collection was carried out, including gathering comments and views from the videos, along with the overall views of the respective channels. Following this, the collected data underwent analysis to identify key patterns and trends. Subsequently, a comparative study of various machine learning (ML) models was performed to evaluate their effectiveness in calculating text sentiment. Each of these steps is discussed in detail in the following sections.

3.1 Data Collection

In the last quarter of 2022, YouTube (https://www.youtube.com) was anonymously and randomly scraped to construct a central repository of channels and videos, serving as the basis for subsequent data collection. The central repository was securely stored in AWS DynamoDB. From January 2023 to December 2023, the YouTube Public API V3 was utilized to systematically gather comments associated with the collected videos, as well as their corresponding view counts.

The data collection process was automated using daily CRON jobs executed on multiple AWS EC2 instances. These CRON jobs were divided into several processes, which were further subdivided into threads using Python. Each thread was responsible for retrieving either the view count of a specific video or the comments for that video on a given day. To minimize redundancy during data collection, a semaphore mechanism was implemented by incorporating flags within the central repository. A visual representation of this process is provided in Figure 3.1.1.

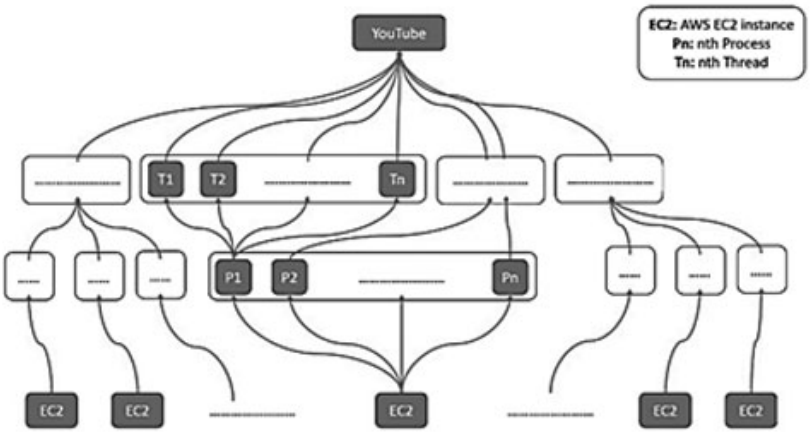


Fig. 3.1.1: Pictorial representation of Multiprocessing & Multithreading used in CRON Jobs for Data Collection

3.2 Data Storage and Normalization

The collected data was initially stored as JSON files in a central AWS S3 bucket. At the end of each day, this data was aggregated and uploaded to AWS DynamoDB for centralized storage. After completing the data collection for the year 2023, the data was downloaded from AWS DynamoDB and stored locally as CSV files. Subsequently, the data was normalized and transferred to a local MySQL database. To minimize

redundancy, the normalized data was organized into multiple relational tables. A snapshot of the MySQL database schema is shown in Figure 3.1.2. Additionally, a table listing the data points, along with their descriptions and data types, is provided in Table 1**.**

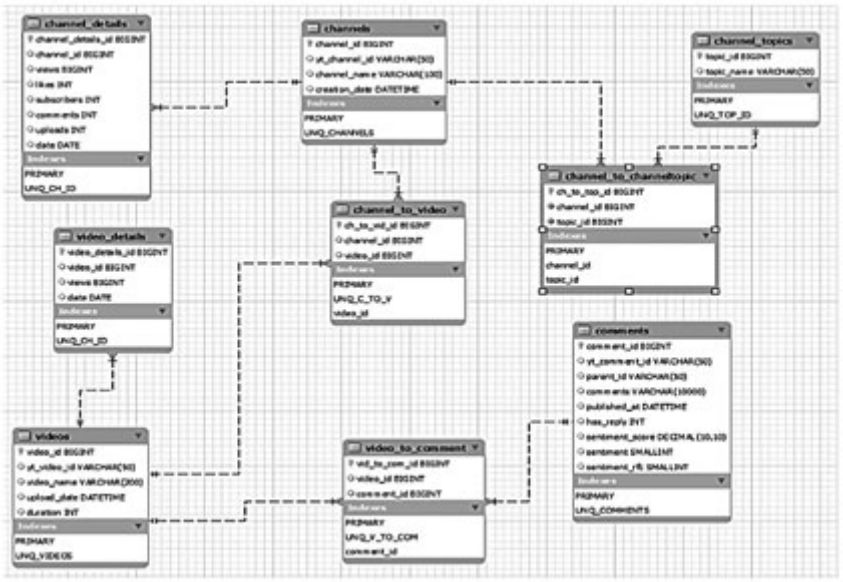


Fig. 3.1.2: MySQL Database Schema

Table 1. Data Description

#	Data Point	Description	Data Type
1.	YouTube Topic	The topic of YouTube Channel	String
2.	YouTube Channel ID	Unique Identifier of YouTube channel	String
3.	YouTube Channel Name	Name of the YouTube Channel	String
4.	Channel Creation Date	Date the channel was created	Date
5.	Channel Views	Total Views of the Channel	Integer
6.	YouTube Video ID	Unique Identifier of YouTube Video	String
7.	Video Name	Title of the YouTube Video	String
8.	Video Upload Date	Date the Video was uploaded	Date
9.	Video Duration	Duration of the video in seconds	Integer
10.	Video Views	Total Views of the Video	Integer
11.	YouTube Comment ID	Unique Identifier of YouTube Comment	String
12.	Comment	Comment on the YouTube Video	String
13.	Comment Publishing Date	Date the comment was made	Date
14.	Parent Comment ID	Identifier of the parent comment	String
15.	Reply	Indicates whether it is the main comment or reply	Boolean

3.3 Text Sentiment Classification

To calculate the sentiment of the text data an open-source LLM was used from Hugginface which was trained on DistilBERT. The sentiment of each comment was predicted and stored in MySQL. Three classes of sentiments were used: *Positive*, *Neutral* and *Negative,* which were encoded as 1, 0, -1, respectively.

3.4 Data Engineering

The data was cleaned to remove or impute null values. The collected comments could be categorized into two classes: *main comments* and *replies* to the main comments. Comments which were replies and whose main comment was not present were dropped as it comprised of only 0.03% of the collected comments. This was done to ensure we have replies only for the main comments that were published in 2023.

The text data was then cleaned and it was converted into lowercase. Then the data was made devoid of punctuation, stop words, HTML tags, URLs, and Special Characters. A new column was added to the comments table indicating whether it is a reply or a comment.

3.5 Data Analysis

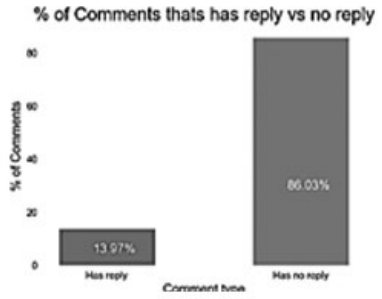
Firstly, the comment sentiments were analyzed to draw insights from them. Then the relationship between the comment sentiments and views on various levels aka, *Videos*, *Channel* and *Topics* were analyzed to draw insights. This was done by exploring the data and then confirming the findings using a statistical test - *One-way ANOVA*. Secondly, a few ML and DL models were compared for text sentiment classification. The models used were *Decision Tree Classifier, Random Forest Classifier, XGBoost Classifier, Simple Sequential CNN model*, *Sequential LSTM model* and a *Transformer* based *Attention model* from *Huggingface* was used. The results and findings are discussed in the next section.

4. Results and Discussions

This section deals with the findings and results of our analysis. Here we will look at the EDA and Statistical test on our data, followed by a comparative study of ML models for Text Sentiment Classification.

4.1 Data Analysis Findings and Results

First, we will look at some of the charts related to the comments and then summarize our findings.



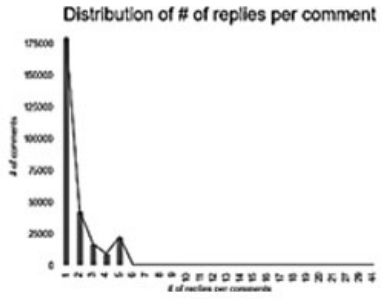


Fig 4.1.1: Distribution of comments on replies *Fig 4.1.2: Distribution of replies per comment*

Figure 4.1.1 above shows us that about 14% of the main comments have replies associated with them, and 86% comments had no replies. We can safely say that these 14% of comments were the most engaging comments. Also, we see that most

comments have between 1-5 replies per comment (Fig. 4.1.2), with 1 reply skewing the comment distribution to the left.

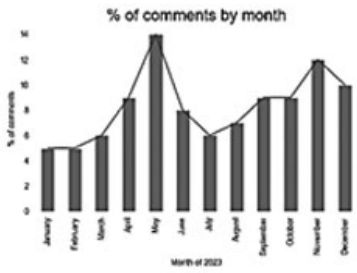


Fig 4.1.3: %age of Comments by Month

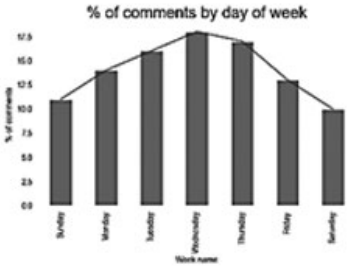


Fig 4.1.4: %age of Comments by Day of Week

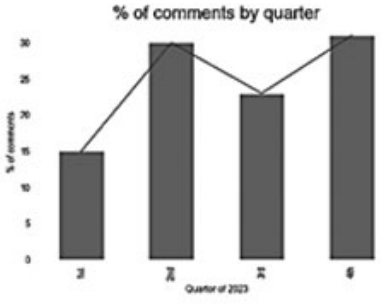


Fig 4.1.5: %age of Comments by Quarter

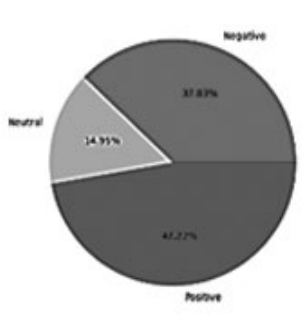
Fig 4.1.6: %age of Comments by Hour of the day

From the figures above, we can say that the 2nd & 4th Quarters of 2023 saw the maximum comments being published in our sample (Fig. 4.1.5). This can be due to people being free to engage in these months because of various seasonal holidays. Breaking them by month (Fig 4.1.3) shows that most comments were published in May 2023, followed by November and December 2023 being the second and third highest months. Again, these are the months that fall under the holiday season. So, content engagement should be high in these months.

Figure 4.1.4 indicates that comments have been made mostly on weekdays rather than on weekends, with Wednesday being the day on which most comments were published, followed by Thursdays and Tuesdays. Figure 4.1.6 shows that most people commented between 1-5 pm, with the peak at 3 pm, and also this trend was declining but prevalent as the day drew further towards the night. These findings indicate that our comments may belong to the students as well as to the working professional classes.

Fig 4.1.7: 1-gram Word Cloud of All comments

Fig 4.1.7 shows us the word cloud of 1-gram. This word cloud indicates the words most used are Love, Thank, One, Know, and Car.



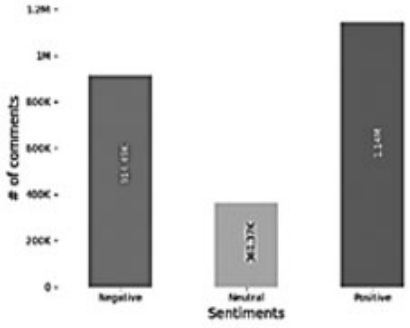


Fig 4.1.8: Comment Sentiment Pie Chart *Fig 4.1.9: Comment Sentiment Bar Chart*

Figures 4.1.8 & 4.1.9 show that most comments there are 20% more positive comments than negative comments. Neutral comments are very less in number, which is even less than the difference between the percentage of positive and negative comments.
Now, the data hierarchy in our dataset is that there are many YouTube topics, each topic has many YouTube channels within them. Each channel has many YouTube videos within them. And we have comments for each video. The diagrammatic representation of this hierarchy is given in Figure 4.1.10.

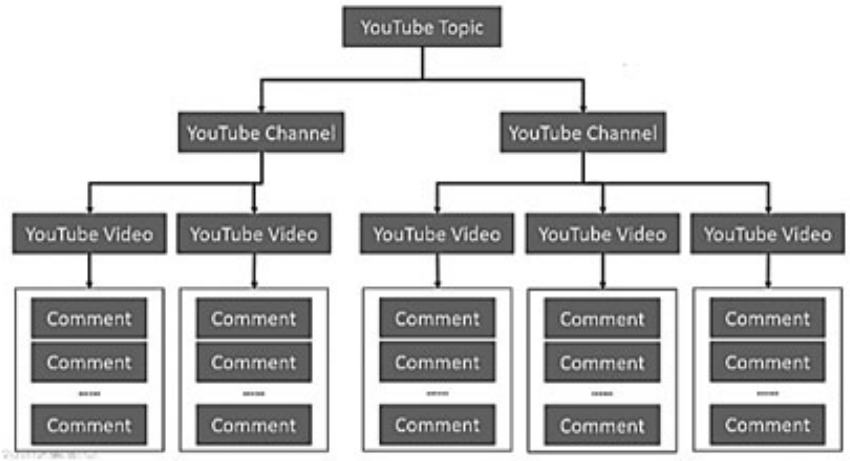


Fig 4.1.10: Data Hierarchy of the Dataset

In order to calculate the sentiment of each level separately (video, channel & topic), we used the simple statistical method of weighted average. The formula for which is given below.

$$Sentiment\ Score = \frac{\sum(sentiment_i * no.of\ comments\ in\ sentiment_i)}{\sum(no.of\ comments)_i}$$

The videos, channels & topics score was then mapped as follows:

- $-1 \leq score \leq -0.25 \rightarrow Negative$
- $-0.25 < score \leq 0.25 \rightarrow Neutral$
- $0.25 \leq score \leq 1 \rightarrow Positive$

Now we will look at the insights drawn between each level and their sentiments.

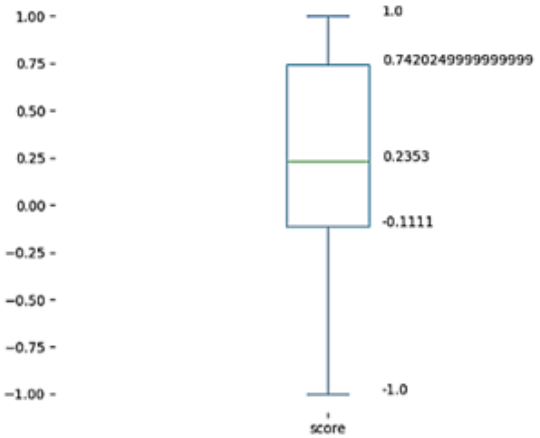
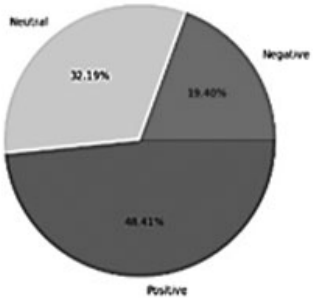


Fig 4.1.11: Video Sentiment Score Boxplot

This boxplot (Fig 4.1.11) indicates that we have 25% data between -1 & -0.1, next 25% data between -0.1111 & 0.2353, next 25% data between 0.2353 & 0.7420 and the last 25% data between 0.7420 & 1. This tells us that about 50% of the data falls between -0.1111 & 0.7420. This would imply that there are more videos with positive than neutral sentiment and even lesser negative sentiment. And these can be confirmed from the pie chart (Fig. 4.1.12) and bar charts (Fig. 4.1.13) below.



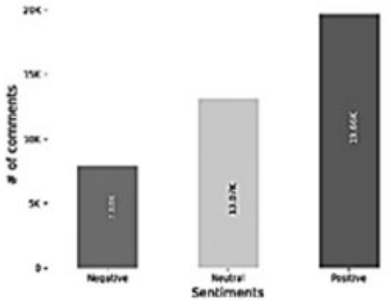


Fig 4.1.12: Video Sentiment Pie Chart *Fig 4.1.13: Video Sentiment Bar Chart*

When we look at the monthly views of videos (Fig. 4.1.14) by the sentiment type, we see that the videos' performance pattern tends to remain the same. Though the videos with positive sentiment outperform the videos with negative sentiment, which in turn outperform videos with neutral sentiment.

Fig 4.1.14: Monthly Video views by Sentiment

Now we will divide the videos into 4 groups by their quartiles formed by *min & q1, q1 and median, median & q3, q3 & max*, and will visualize the sentiment distribution by these groups. Figure 4.1.15 shows that all the quadrant has the same distribution of sentiments, with positive outperforming the negative, which in turn outperforms the neutral ones.

On looking at the video views separated by sentiment & view quadrants (Fig 4.1.16), we see that there is a difference in performance among the sentiment groups but there is not much difference among the quadrants.

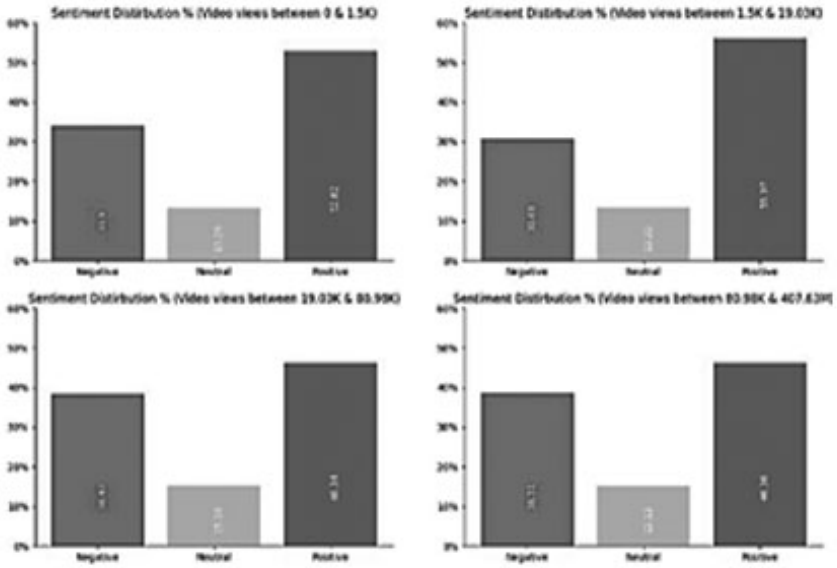


Fig 4.1.15: Video Sentiment by Video Views Quartiles

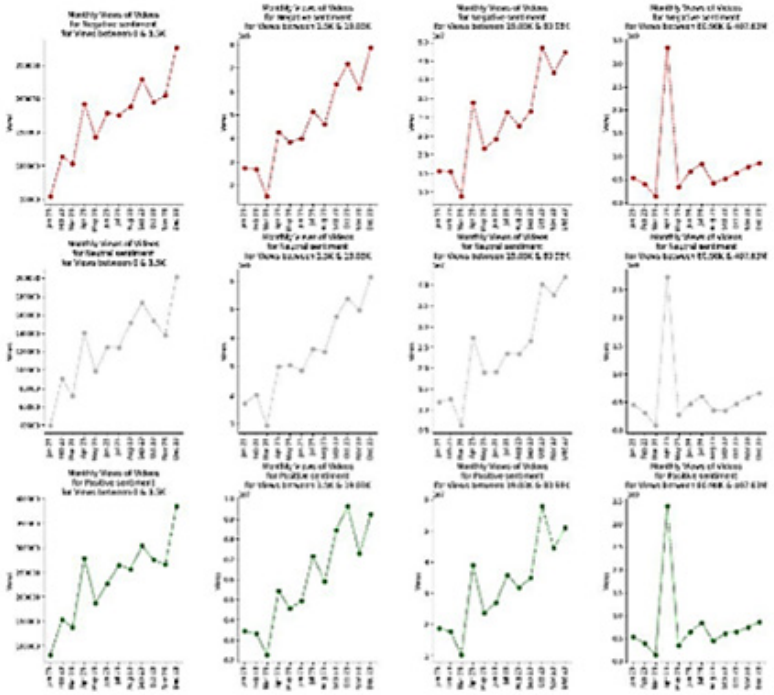


Fig 4.1.16: Monthly Video views by Sentiment and Views Quadrant

This indicates that the video sentiments from comments are not the best indicator of a video's performance with respect to views.

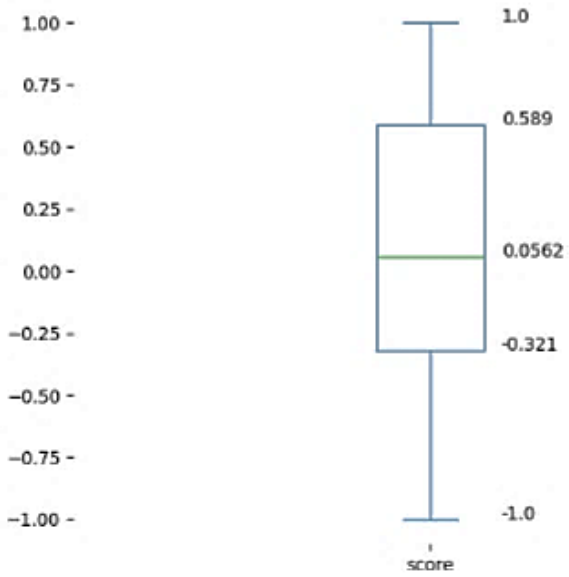


Fig 4.1.17: Channel Sentiment Score Boxplot

This boxplot (Fig 4.1.17) indicates that we have data 25% data between -1 & -0.321, next 25% data between -0.321 & 0.0562, next 25% data between 0.0562 & 0.589 and the last 25% data between 0.589 & 1. This tells us that about 50% of the data falls between -0.321 & 0.589. This would imply that there are more channels with positive than neutral sentiment, and even fewer with negative sentiment. And these can be confirmed from the pie chart (Fig 4.1.18) and bar chart (Fig 4.1.19) below.

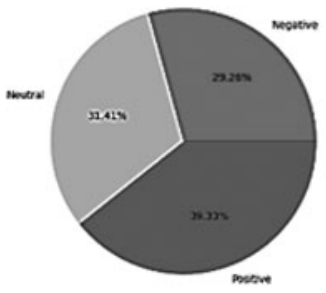


Fig 4.1.18: Channel Sentiment Pie Chart

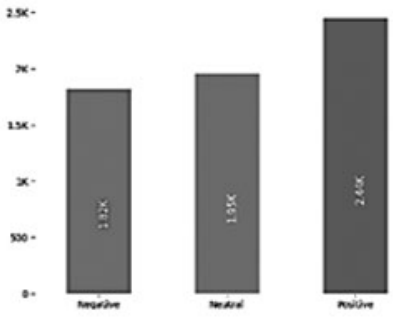


Fig 4.1.19: Channel Sentiment Bar Chart

When we look at the monthly views of channels (Fig. 4.1.20) by the sentiment type, we see that the performing pattern of channels is not the same for the various classes of sentiments. The channels belonging to the neutral class perform better overall than the other two. Channels belonging to the positive class perform comparatively lesser than the other two.

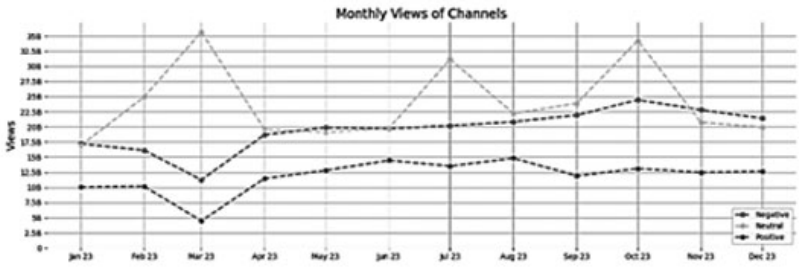


Fig 4.1.20: Monthly Channel views by Sentiment

Now we will divide the channels into 4 groups by their quartiles formed by *min & q1, q1 and median, median & q3, q3 & max*, and will visualize the sentiment distribution by these groups.

Figure 4.1.21 shows that all the quadrants do not have the same distribution of sentiments.

Positives are high in the *first* and *second* quadrant, the neutral is higher in the *third* and the negative is higher in the *fourth* quadrant. This would mean that performance is not dependent on the sentiment class of the channel.

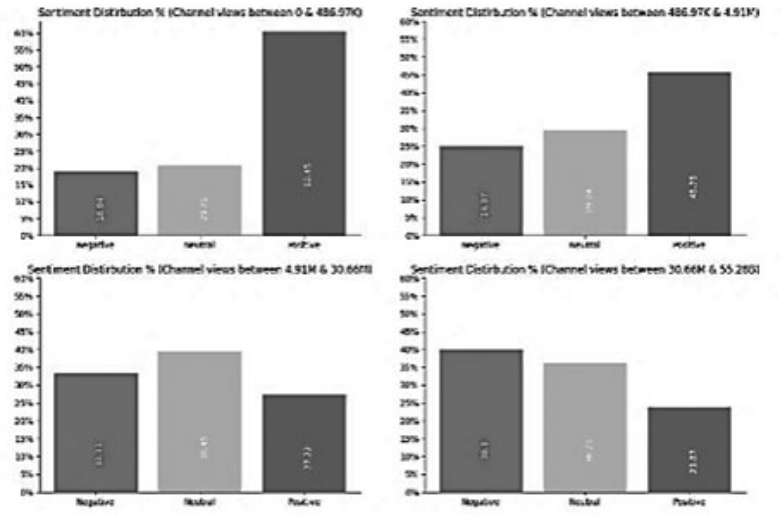


Fig 4.1.21: Channel Sentiment by Channel Views Quartiles

On looking at the channel views separated by sentiment & view quadrants (Fig. 4.1.22), we see that there is a difference in performance among the sentiment groups, but there is not much difference among the quadrants.

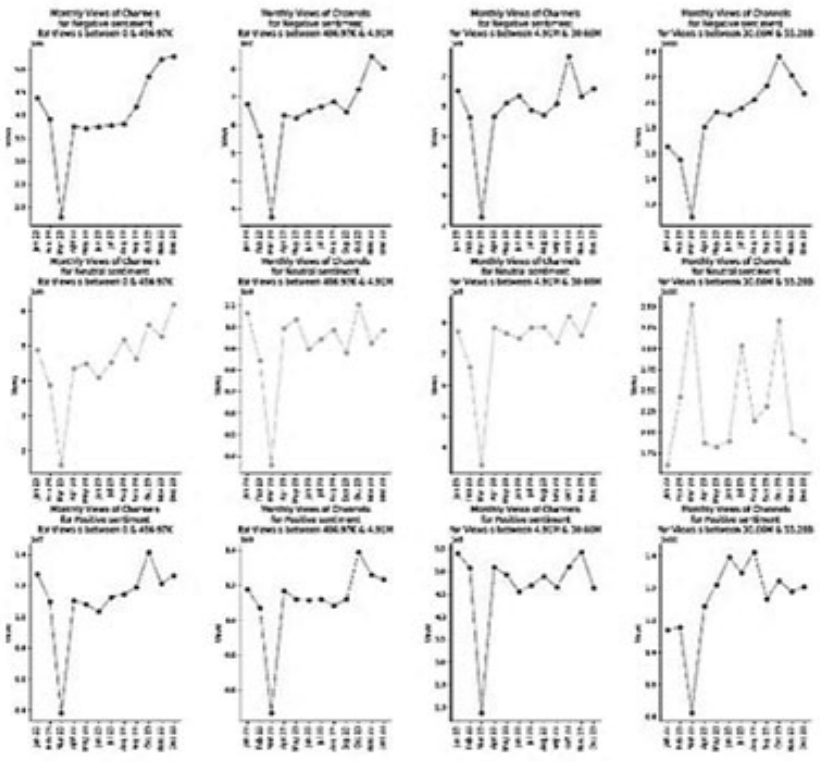


Fig 4.1.22: Monthly Channel views by Sentiment and Views Quadrant

This indicates that there *is a difference* in performance of the sentiment classes, but is not an indicator of a *channel's* performance with respect to views.

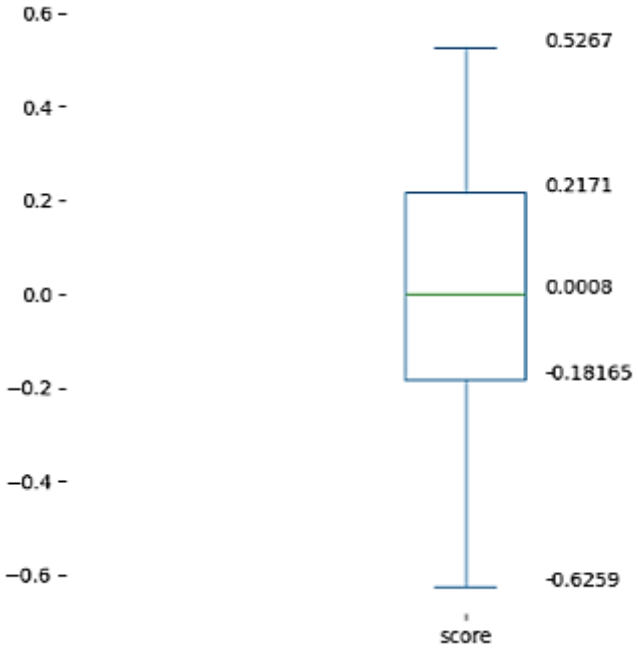
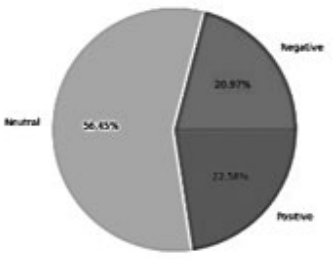


Fig 4.1.23: Topic Sentiment Score Boxplot

This boxplot (Fig 4.1.23) indicates that we have data 25% data between -0.6259 & -0.18165, next 25% data between -0. 18165 & 0.0008, next 25% data between 0. 0008 & 0.2171and the last 25% data between 0. 2171 & 5267. This tells us that about 50% of the data falls between - 0.18165 & 0.2717. This would imply that there are more topics with neutral sentiment than positive sentiment and negative sentiment. On examining Fig 4.1.24 & Fig 4.1.25 we can say that there are more topics that belong to the neutral class sentiment than the other two. In fact, % of topics belonging to neutral class is greater than the sum of % of comments belonging to the positive & negative classes.



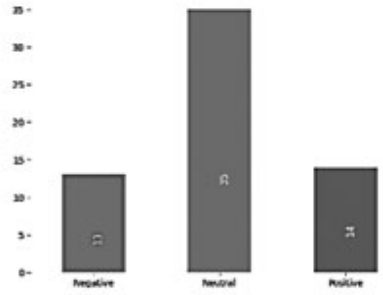


Fig 4.1.24: Topic Sentiment Pie Chart *Fig 4.1.25: Topic Sentiment Bar Chart*

When we look at the monthly views of topics (Fig 4.1.26) by the sentiment type, we see that the performing pattern of topics is not same for the various classes of sentiments. The topics belonging to the neutral class outperform the other two. Topics belonging to the negative class perform comparatively better than the ones belonging to the positive class.

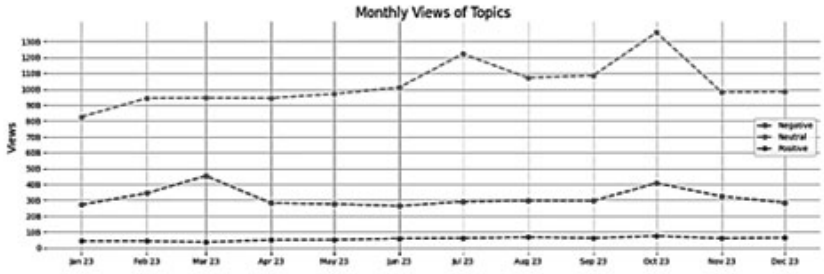


Fig 4.1.26: Monthly Topic views by Sentiment

Now we will divide the topics into 4 groups by their quartiles formed by *min & q1, q1 and median, median & q3, q3 & max*, and will visualize the sentiment distribution by these groups. Figure 4.1.27 shows that all the quadrants do not have the same distribution of sentiments. Negatives are high in the first quadrant, neutral is higher in all the other quadrants. This would mean that performance is not dependent on the sentiment class of the topic.

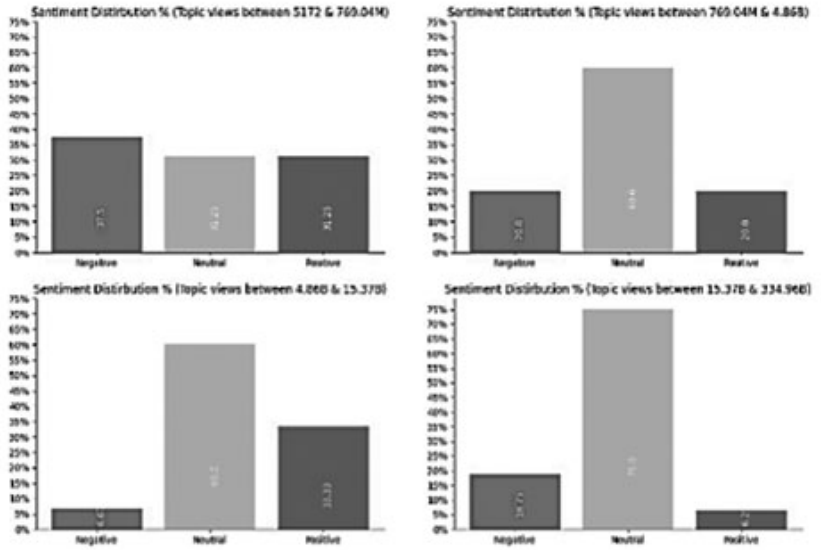


Fig 4.1.27: Topic Sentiment by Topic Views Quartiles

On looking at the topic's views separated by sentiment & view quadrants (Fig 4.1.28), we see that there is a difference in performance among the sentiment groups but there is not much difference among the quadrants.

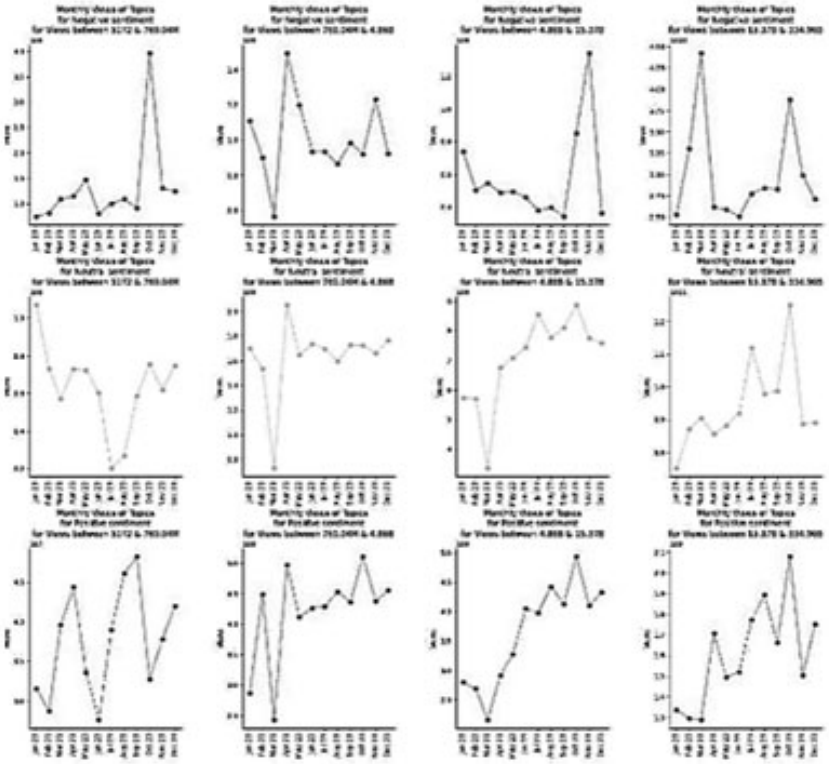


Fig 4.1.28: Monthly Topic views by Sentiment and Views Quadrant

This indicates that there is a difference in performance of the sentiment classes, but it is not the best indicator of a topic's performance with respect to views.
We also performed the following statistical tests to confirm our findings. The results of which are tabulated below.

Level	Test Statistic	P-value	Inference
Videos	4.538344963413858	0.018146553063464	p-value < 0.05 => Rejecting Null Hypothesis This implies that there is a difference in performance among the groups.
Channels	23.20820302701746	5.09285395889497*10-7	
Topics	385.31733981251443	1.324377995578209*10-23	

The results above indicate that we have sufficient proof to say that there exists a difference in the performance among the three different sentiment classes for *Videos*, *Channels* & *Topics*.

4.2 Model Comparison

The computational models for text sentiment classification were trained on a dataset of 30,000 data points, with 10,000 data points belonging to each class. This would ensure a balanced training dataset. The results are given in the table below:

Table 3: Model Performance Table

Model	Training Accuracy	Testing Accuracy
Random Forest Classifier	57.62%	53.23%
Decision Tree Classifier	59.26%	51.35%
XGBoost Classifier	53.62%	48.60%
CNN Classifier	96.69%	78.12%
LSTM Classifier	33.29%	33.58%
Pre-Trained LLM	*100.00%*	*98.33%*

From the table above it can be inferred that Attention-based Transformer models outperform other ML models.
The classification Reports of the models are given below.

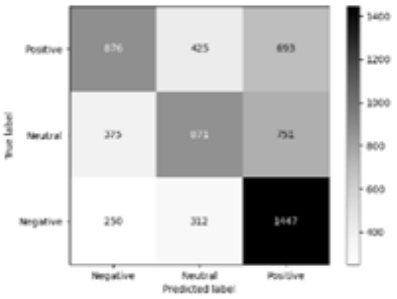


Fig 4.1.29: Confusion Matrix of Random Forest

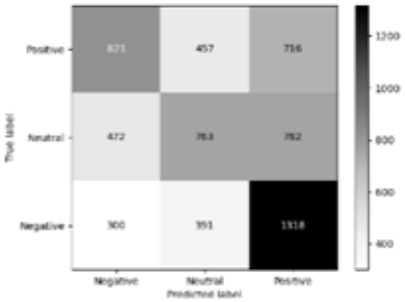


Fig 4.1.30: Confusion Matrix of Decision Tree

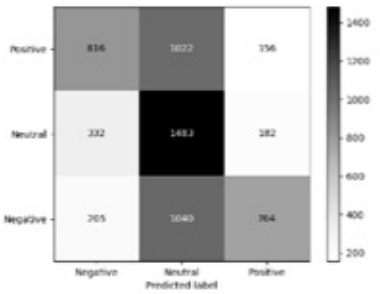


Fig 4.1.31: Confusion Matrix of XGBoost

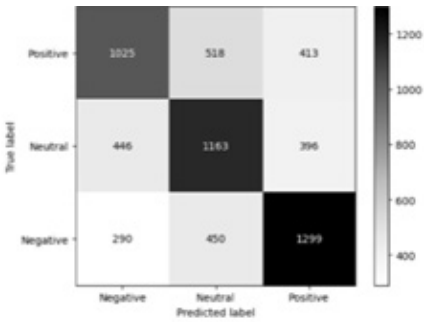


Fig 4.1.32: Confusion Matrix of CNN

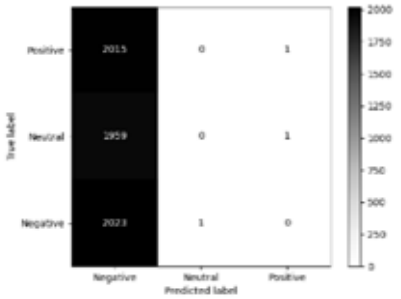


Fig 4.1.33: Confusion Matrix of RNN

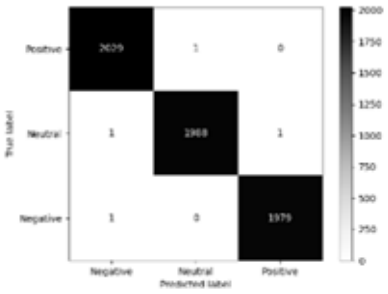


Fig 4.1.34: Confusion Matrix of LLM

5. Conclusion and Future Scope

From our analysis, we see that the *performance of videos, channels, and topics varies differently at each level*. But the *sentiments* are *not* the best indicator of the performance of a given level. This is because view is one of the KPI and there are other metrics that affect the performance as well. We also see that an LLM pre-trained for text sentiment classification outperforms any other ML model. This happens because the capability of LLMs to remember the context within the sentence is much higher than other models. Future works that can be done may include using a different score calculation formula for sentiment of the various levels. Also, other metrics related to YouTube can be explored.

♣ References

1. Athar, A. (2014). Sentiment analysis of scientific citations. *Computational Linguistics, 40*(3), 587–592. https://doi.org/10.1162/COLI_a_00190
2. Athar, A., & Teufel, S. (2012, July). Detection of implicit citations for sentiment detection. In *Proceedings of the Workshop on Detecting Structure in Scholarly Discourse* (pp. 18–26).
3. Balog, K., Azzopardi, L., & de Rijke, M. (2013). A system for analyzing Twitter sentiment in the smartphone market. *Journal of Information Retrieval Studies, 18*(3), 45–63.
4. Balog, K., Kenter, T., & Voskarides, N. (2018). Demographic-aware sentiment analysis using Twitter data. *Journal of Computational Social Science, 2*(1), 112–129.
5. Frank, E., Trigg, L., Holmes, G., & Witten, I. H. (2003). Naive Bayes for regression. *Machine Learning, 41*(1), 5–25. https://doi.org/10.1023/A:1024442310105
6. Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter sentiment analysis: The good, the bad and the OMG! In *Proceedings of the Fifth International Workshop on Semantic Evaluation* (pp. 538–541).
7. Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter sentiment analysis: The good, the bad and the OMG! In *Proceedings of the Fifth International Conference on Weblogs and Social Media* (pp. 538–541). AAAI.
8. Kowcika, A., Sharma, R., & Patel, S. (2019). A sentiment analysis system for the smartphone market using Twitter data. *Journal of Data Analytics and Applications, 5*(2), 112–120.

9. Kramer, A. D. I., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences, 111*(24), 8788–8790.
10. Maynard, D., Bontcheva, K., & Rout, D. (2014). Challenges in developing opinion mining tools for social media. *Proceedings of the International Conference on Weblogs and Social Media (ICWSM), 15*(1), 15–21.
11. Mishne, G. (2006). Experiments with blog sentiment analysis. In *Proceedings of the 1st Workshop on Stylistic Analysis of Text for Information Access*.
12. Mishne, G., & de Rijke, M. (2005). Capturing global mood levels using blog posts. In *Proceedings of the AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs* (pp. 145–152).
13. Parvathy, G., & Bindhu, J. S. (2016). A probabilistic generative model for mining cybercriminal networks from online social media: A review.
14. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*.
15. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval, 2*(1–2), 1–135.
16. Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing* (Vol. 10, pp. 79–86). Association for Computational Linguistics.
17. Poria, S., Cambria, E., Gelbukh, A., Bisio, F., & Hussain, A. (2015). Sentiment data flow analysis by means of dynamic linguistic patterns.
18. Qazvinian, V., & Radev, D. R. (2010, July). Identifying non-explicit citing sentences for citation-based summarization. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics* (pp. 555–564). Association for Computational Linguistics.
19. Riboni, D. (2004). Feature selection for website classification. In *Proceedings of the 13th International World Wide Web Conference* (pp. 123–134).
20. Saif, H., He, Y., & Alani, H. (2012, November). Semantic sentiment analysis of Twitter. In *International Semantic Web Conference* (pp. 508–524). Springer, Berlin, Heidelberg.
21. Shree, S., & Brolin, J. (2019). Sentiment analysis of YouTube comments using an unsupervised lexicon-based approach. *International Journal of Social Media Studies, 12*(3), 45–57.
22. Siersdorfer, S., Chelaru, S., Nejdl, W., & San Pedro, J. (2010). How useful are your comments?: Analyzing and predicting YouTube comments and comment ratings. In *Proceedings of the 19th International Conference on World Wide Web* (pp. 891–900).
23. Socher, R. (2016). Deep learning for sentiment analysis—Invited talk. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media Analysis*.

24. Sobhani, P., Mohammad, S., & Kiritchenko, S. (2016). Detecting stance in tweets and analyzing its interaction with sentiment. In *Proceedings of the 5th Joint Conference on Lexical and Computational Semantics*.
25. Turney, P. D., & Mohammad, S. M. (2014). Experiments with three approaches to recognizing lexical entailment.