Group Two

Assignment Five

Presented to Dr. Andy Asare; Khizer Kamran

# UNIVERSITY OF CALGARY

Anika Javed, Jody Nguyen, Matthew Haddad, Sayed Subhan Shah Sadat, Sofiri Afonya

Introduction to Actionable Data Visualizations

DATA 605

DataThon 5

# Contents

# Introduction

Within this age of social media where everyone has a different algorithm designed for them, it is common for there to be specific moments that reach multiple social feeds, a piece of entertainment that is viewed by a wide demographic of people due to its many factors that bridge the gaps within everyone's interests. This project dives into the roots of what creates a viral content piece and will use the infamous video of Will Smith slapping Chris Rock during the Oscars as the main example as to how viral media can capture many people's attention in a brief period of time. We will analyze public sentiments surrounding the viral post using a range of analytical tools that will help to visualize the overall sentiment, public's subjective and object engagement with the post. And illustrate how the post spread through different online communities and its impact on public discourse.

The project also assesses the virality of this event and its effects on the public perception of integrity related to the Oscars and the actors involved. While the popularity of these events can result in increased attention and audience size for the next event, it can also lower the pedestal that Hollywood actors have tried to build for so long, which would call for ways of damage control to bounce back from the negative attention. We will provide recommendations for the Academy and the parties involved on how to address and resolve the issue. But if the sentiment is mostly positive, we will propose ways for the Academy to capitalize on the viral post, to enhance their business performance through engagement.

# Guiding Questions and Objectives

Question 1 (Sofiri): In this video, which are the most frequent words that were used in the comment section?

Objective: Try to make conclusions about the video based on the main buzzwords that were used in the comment section through a visualization.

Data Visualization and Findings

One of the main visualizations we will look at to add to the interpretations of the report is the following word cloud:

We can use these results to make inferences as to how the public perceived these videos in a neutral lens. We will dismiss ideas of subjectivity and objectivity and purely try to make conclusions on the source based on the use of specific words, and the popularity of usage amongst them. Without even knowing the source of this word cloud, we can clearly see that the main focuses on "who's" in the video are Will Smith, Jada, and Chris Rock. The "how" or actions can be categorized from words such as slap and joke, and the "why" can be seen through values such as respect and wife. Now looking through the lens of someone who has watched the video, these words still give the same amount of sense. Knowing that at the **Oscars**, a joke by the **comedian Chris Rock** that was made about **Jada**, **Will Smith's wife**, resulted in a **slap** that

was motivated by having **respect** for his wife. The opinions on this matter can be summed up through words such as "deserved," "funny," "scripted" and "violence."

There is the group of people that thought that the slap was deserved, along with the group of people who think the overall event was funny, planned, or was unnecessary violence. After the slap Will Smith constantly says, "keep my wife's name out of your mouth," which can be seen within the word cloud into separated words.

# Question 2 - What are the primary emotions expressed in online discussions about the incident where Will Smith slapped Chris Rock at the Oscars?

## Objective

The objective is to determine the range of emotions expressed in the comments, such as anger, surprise, humor, support, criticism. A dendrogram is created to visualize the hierarchical clustering of comments based on their sentiment scores.

## Data Visualization and Findings



Dendrogram of Topic Polarity
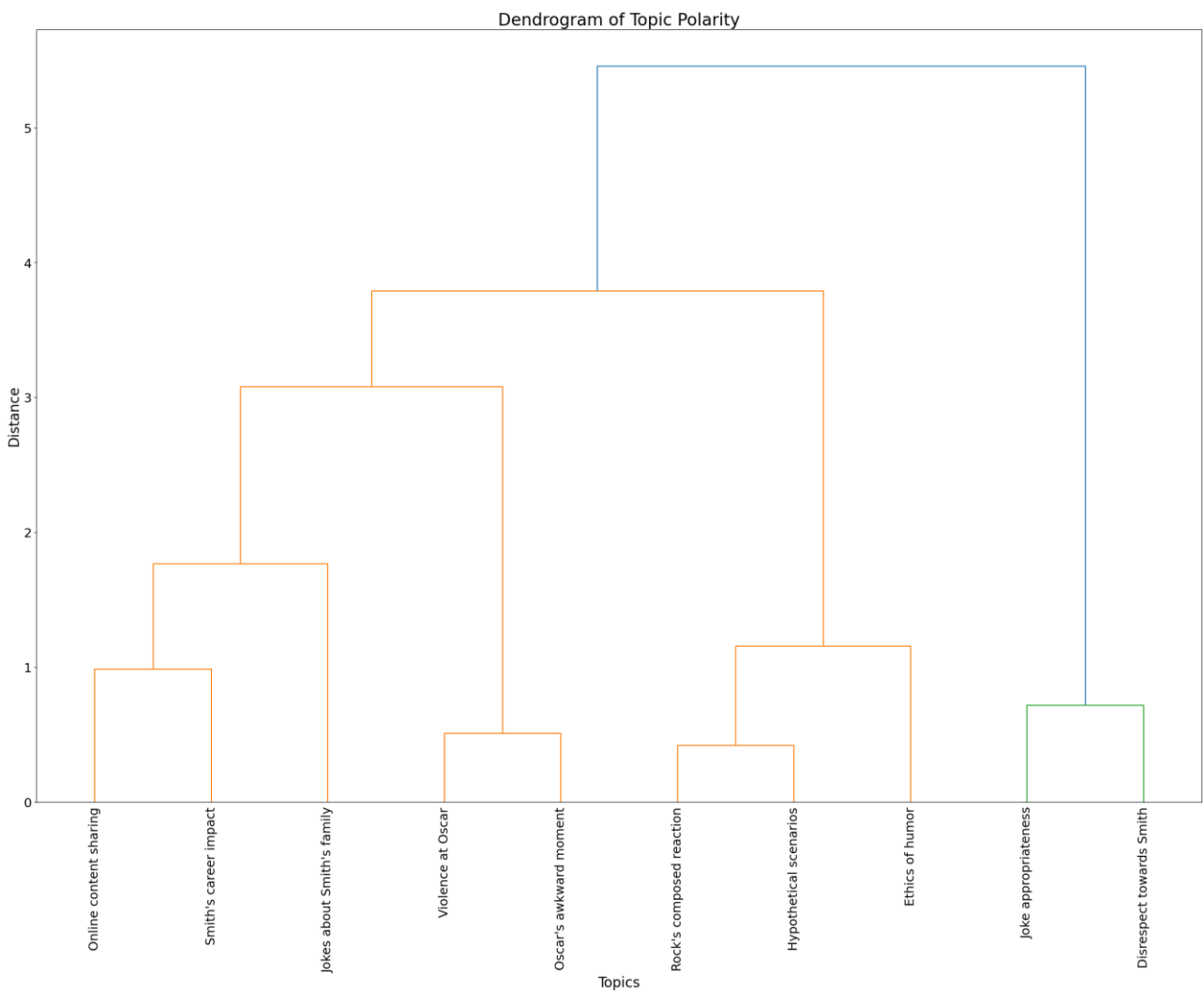
Firstly, the first primary cluster's sub-cluster, which groups "Online content sharing" and "Smith's career impact," suggests discussions around public perception and Smith Career's impact.

The next level adds connected Oscar's awkward moment when Will Smith slapped Chris Rock comments likely discussing the direct incident and its immediate awkwardness during the Oscar ceremony.

Another sub-cluster which encompasses "Hypothetical scenarios," "Rock's composed reaction," and "Ethics of humor." This indicates that discussions around hypothetical alternatives to the incident, Rock's reaction, and the broader ethical considerations of humor in public settings share thematic similarities.

Secondly, the secondary cluster in Green, which includes "Joke appropriateness" and "Disrespect towards Smith," indicates that these topics are closely related. This suggests acceptability of humor, or the appropriateness of the joke could be closely related to Smith's reactions against the joke that Chris made about his wife Jada.

Based on the analysis of the dendrogram and the identified clusters, the primary emotions expressed in the online discussions regarding the incident where Will Smith slapped Chris Rock at the Oscars can be categorized into several key areas:

- Shock and Surprise: This emotion is primarily associated with the cluster including "Oscar's awkward moment"

- Disapproval and Criticism: These sentiments are evident in discussions about "Disrespect towards Smith" and the "Ethics of humor."

- Concern and Empathy: Within the cluster of "Smith's career impact" and "Online content sharing," there is an underlying concern for the personal and professional career for Smith stemming from the incident.

In conclusion, this refined analysis provides a clearer understanding of how various topics related to the Will Smith and Chris Rock incident at the Oscars are interconnected. The accurate clustering reflects deeper insights into how people discuss and relate different aspects of the incident, from ethical considerations to personal impacts and public reactions.

# Question 3 (Matthew Haddad): How do the different topics differ from one another?

Objective: Create visuals that clearly show patterns across 10 types of YouTube comments.

Data Visualization and Findings

The 10 topics that were grouped in an LDA analysis are:

1. **Jokes about Smith family:** Jokes made at the expense of the family of Will Smith

2. **Smith career impact:**  Will Smiths actions will affect his future career

3. **Disrespect towards Smith:** The joke made about Will Smiths Wife and how it was disrespectful

4. **Oscar awkward moment:** How awkward the slapping of Will Smith was

5. **Online content sharing:** How the moment is and will be shared currently and, in the future

6. **Joke appropriateness:** How appropriate the joke Chris Rock made was

7. **Hypothetical scenarios:** What the commentors would do in Wills position

8. **Ethics of humor:** If Chris Rocks joke was ethical or not

9. **Rock composed reaction:** How composed Chris Rock was after getting slapped and yelled at

10. **Violence at Oscar:** How Will Smith slapped Chris Rock on live television at the Oscars

To help visualize the topic people were most interested in talking about, a bar chart that displays the count of each type of comment was made:

**Count of Topics**
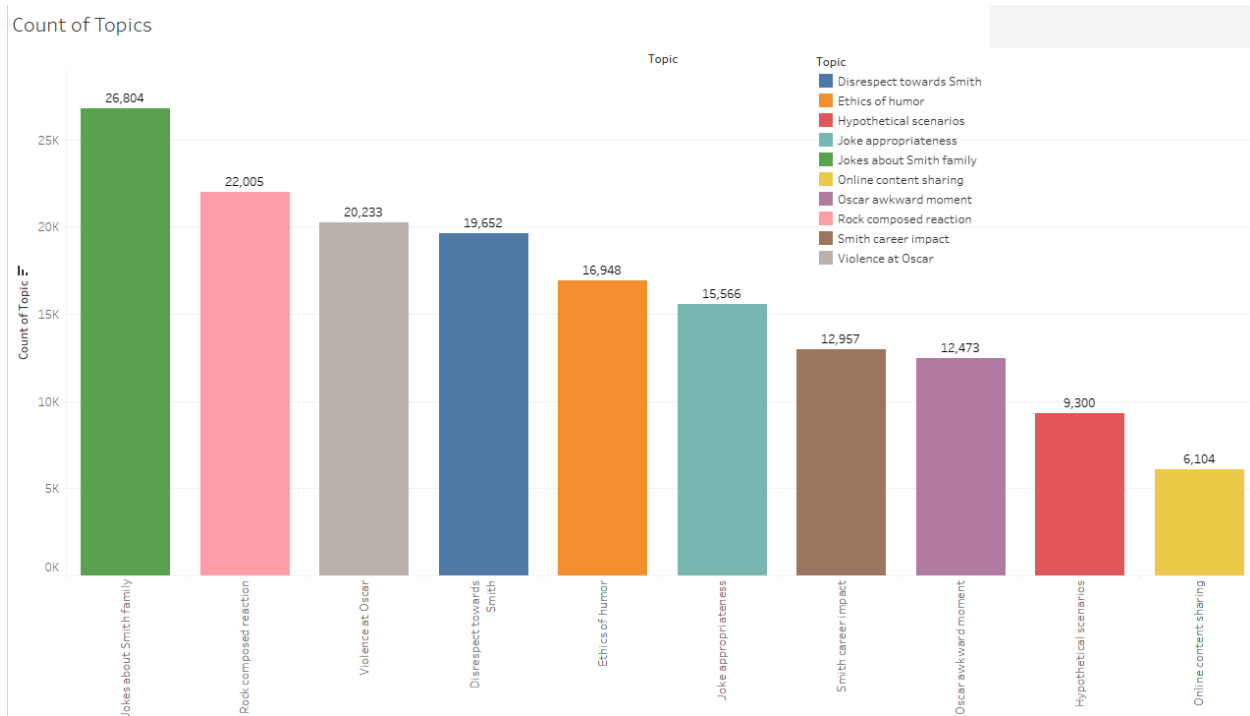
Here we can clearly see that Jokes about the Smith family were the most talked about at 26804 comments. We can see however, that the differences between each successive rank in topic count decreases by a small margin. This can be seen with *Rock composed reaction* in a close second place with 22005 counts. The topic with the least number of counts is *Online content sharing*. This is due to people wanting to comment more on the situation at hand more than how it will be shared.

To further this investigation, we want to know what topics people agreed the most on. We can assume that the topic with the most likes is the most agreeable, we will display a chart to show this:

## Likes vs Topics
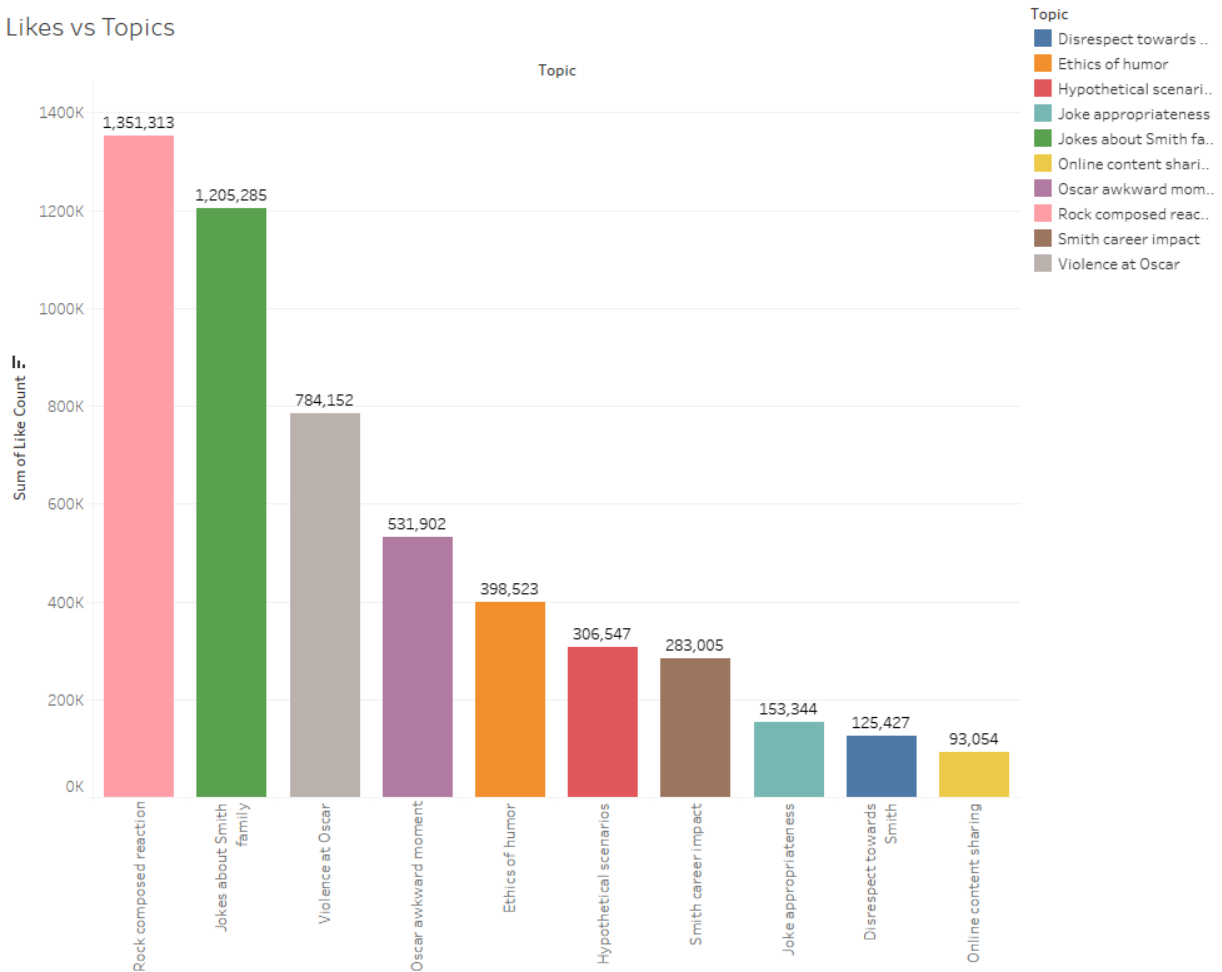


Here we can see that *Rock composed reaction* was the most liked topic with 1351313 likes. In a close second, we see that *jokes about the Smith family* with 1205285 likes takes this place. With two main parties: Will and Chris, we can see that people agree that Chris was composed, and that Will smith made a mistake. *Disrespect towards Smith* is in the penultimate spot with 125427 likes, indicating that few people agreed that Will was disrespected.

To gauge whether people were subjective or objective in their commenting. A Subjective analysis was conducted displaying the comment counts of fact and opinion per topic:

Fact or Opinion
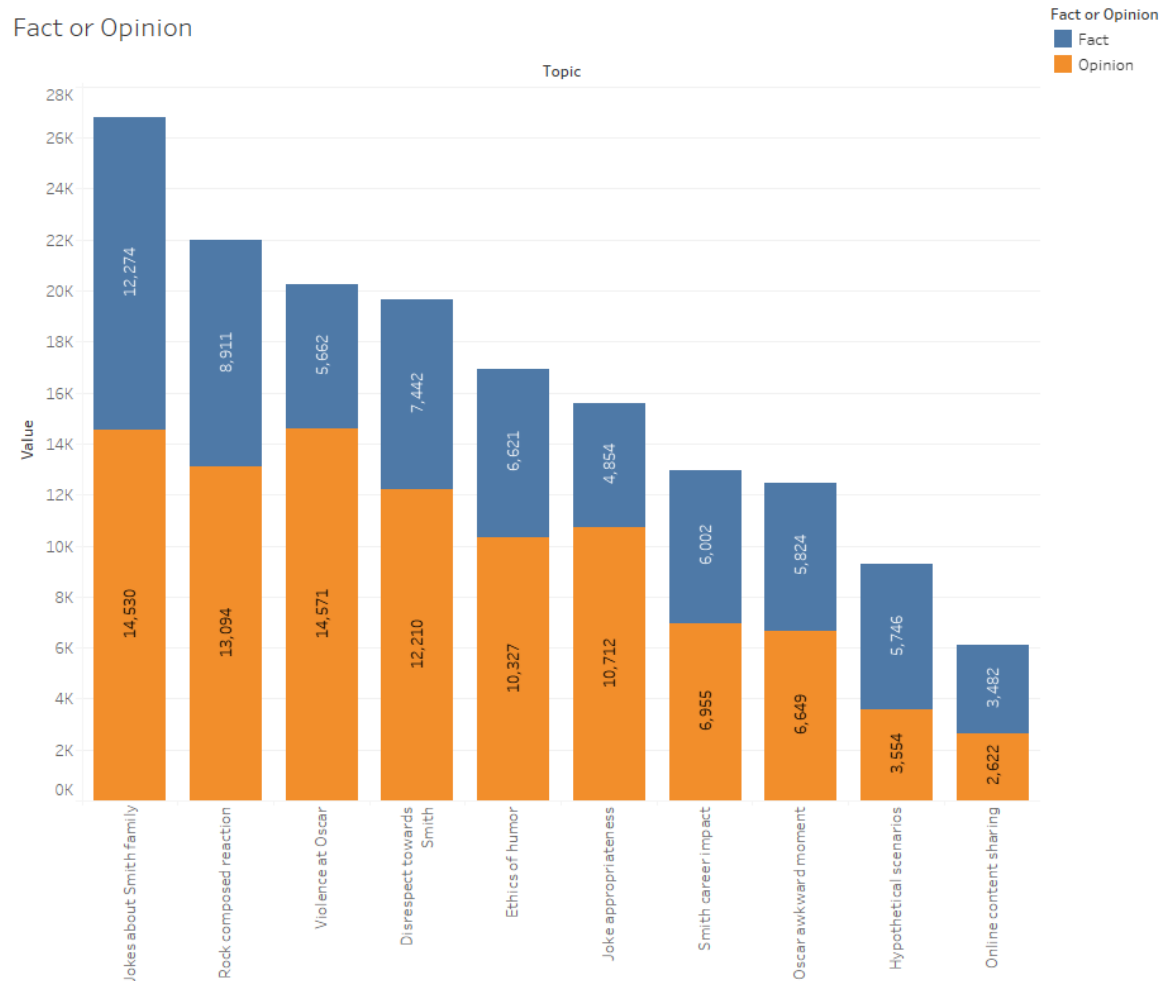
Here we see a general trend of more opinion than fact. However, in topics: *Hypothetical scenarios* and *Online content sharing, we* see a distribution of more fact than opinion. An explanation for this could be that when people are stating a hypothetical, they need to do so based off what they already know, which are the facts of the scenario. Furthermore, when talking about how the video will be shared, it must be talked about with how people have seen viral videos become viral, facts. We also notice that *Violence at Oscar* has the largest percentage of opinion-based comments. Is result is contradictory as the fact that Will did indeed slap Chirs. This result could be explained as people could have different interpretations of the events that took place and whether the slap could be classified as violence.

Finally, we want to look at the sentimentality of each topic and how the comments differ in, positivity, negativity, and neutrality:
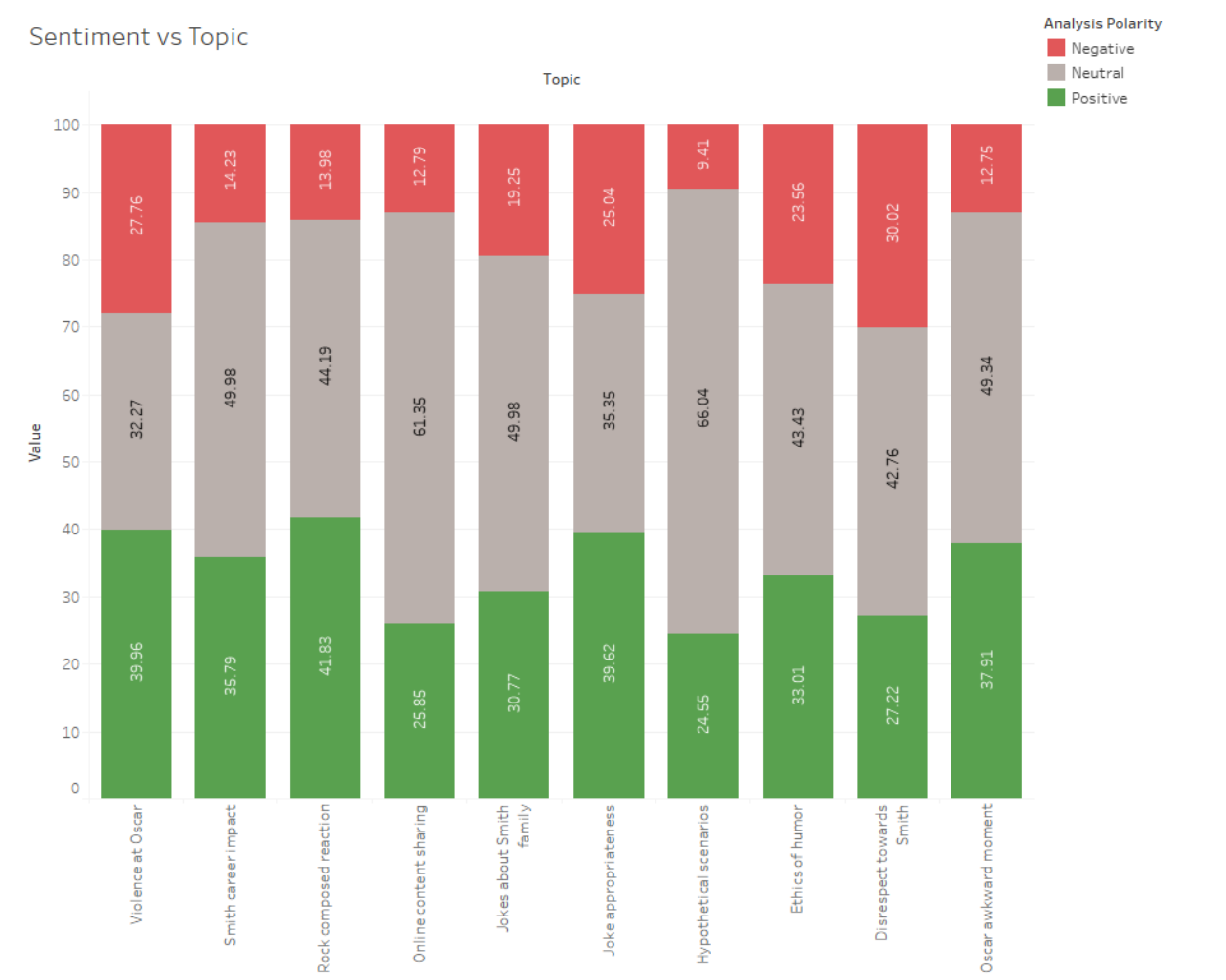


Here we see that the most positive topic is *Rock composed reaction at 41.83%.* This makes sense as most people were complimenting Chris on his composure for hosting the Oscars on live television while having to deal with getting slapped in the face. The most neutral topic is *Hypothetical scenarios* at 66.04%. Discussing hypothetical scenarios often requires critical
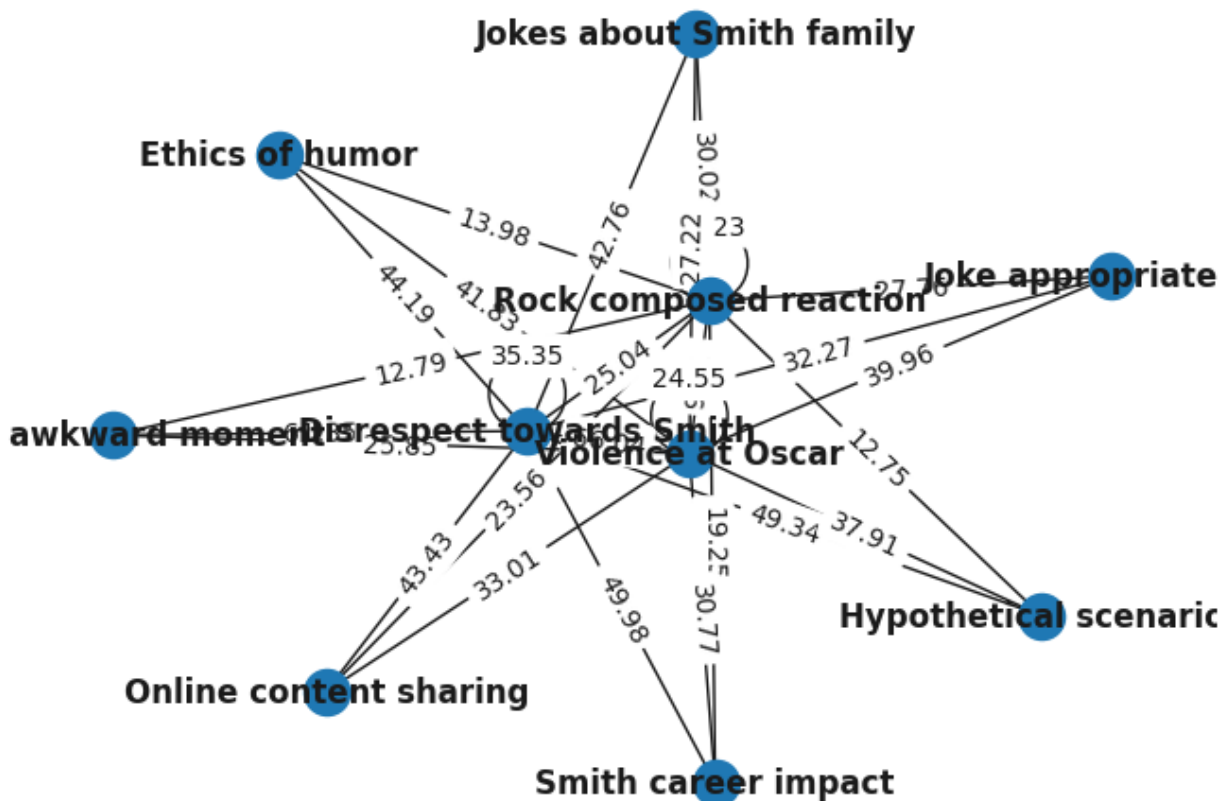
thinking and intellectual engagement. Commenters may focus on analyzing and discussing potential implications, leading to a greater emphasis on reasoned and balanced commentary. The most negative topic is *Disrespect towards Smith* at 30.02%. Comments about "Disrespect towards Smith" might harbor the most negative sentiment, given the emotionally charged nature of the topic. Discussions on perceived disrespect may evoke strong reactions from Will Smith's supporters, resulting in a higher count of negative comments expressing frustration or dissatisfaction.

## Question 4: (Sayed Subhan Shah Sadat) How do the different discernible topics from the data relate with each other?

Objective: Analyzing the relations of the topics extracted from the data.

Data Visualization and Findings:

Initially, it is important to note that to provide an answer for the fourth question, the comments data had to be analyzed through the Latent Dirichlet Allocation modelling method. Afterwards, a network graph was generated using the polarity matrix for the topics calculated earlier with other python libraries. Lastly, each node/topic was ranked in terms of Centrality and Betweenness which allowed for the visualization of a graph illustrating the rank of each topic with respect to one another. The following visualization was generated through the steps mentioned above:

**Jokes about Smith family**

**Ethics of humor**

**Joke appropriate**

**Rock composed reaction**

**awkward moment** **Disrespect towards Smith** **Violence at Oscar**

**Hypothetical scenario**

**Online content sharing**

**Smith career impact**

13.98  42.76  27.22  23  27
44.19  41.83  35.35  25.04  24.55  32.27  39.96
12.79  12.75
25.85  49.34  37.91
43.43  23.56  19.25  30.77
33.01  49.98

As it can be seen in the network graph above, three topics stand out from the rest in terms of centrality and betweenness, namely:

- Rock composed reaction

- Disrespect towards Smith

- Violence at Oscar

It is arguable that these three topics gathered from the dataset capture the essence of the infamous event that transpired; At its core, the event was a violent action that took place at the Oscar's which was caused by disrespect towards Will Smith after which Chris Rock reacted by

maintaining his composure. The calculated centrality and betweenness values for the nodes are as follows:

**Degree Centrality:**

Rock composed reaction: 1.222222222222222

Disrespect towards Smith: 1.222222222222222

Violence at Oscar: 1.222222222222222

Oscar awkward moment: 0.3333333333333333

Joke appropriateness: 0.3333333333333333

Ethics of humor: 0.3333333333333333

Hypothetical scenarios: 0.3333333333333333

Smith career impact: 0.3333333333333333

Online content sharing: 0.3333333333333333

Jokes about Smith family: 0.3333333333333333

**Betweenness Centrality:**

Rock composed reaction: 0.19444444444444442

Disrespect towards Smith: 0.19444444444444442

Violence at Oscar: 0.19444444444444442

Oscar awkward moment: 0.0

Joke appropriateness: 0.0

Ethics of humor: 0.0

Hypothetical scenarios: 0.0

Smith career impact: 0.0

Online content sharing: 0.0

Jokes about Smith family: 0.0

The relationship between the topics that did not score highly with respect to centrality and betweenness can be well explained with the topics that did see themselves ranked highly regarding the two metrics. For instance, "Online Content Sharing" and "Joke Appropriate" nodes are linked with each other through "Violence at Oscar" and their relationship can be explained in the following manner. "Online Content Sharing" is the virality of the event on social media platforms such as YouTube and Instagram and this event went viral due to an act of violence at the Oscar's which itself was a reaction to a perceived inappropriate joke. Such interpretations for the rest of the peripheral nodes can be readily made with consideration of the central topics.

## Question 5 (Anika Javed) Sentiment Analysis

## What was the sentiment, polarity, and subjectivity of the comments on the Youtube Video?

### Objective

- Conduct polarity and subjectivity analysis using a sentiment analysis library TextBlob.
- Calculate the overall sentiment polarity and subjectivity scores for the comments.

## Data Visualization and Findings

We present an analysis of YouTube comments using sentiment analysis techniques. The aim is to identify and categorize comments based on their sentiment polarity—positive, negative, or neutral—and to determine the most subjective words associated with each sentiment category. The data was processed using the TextBlob library in Python, which provided sentiment scores for each comment.
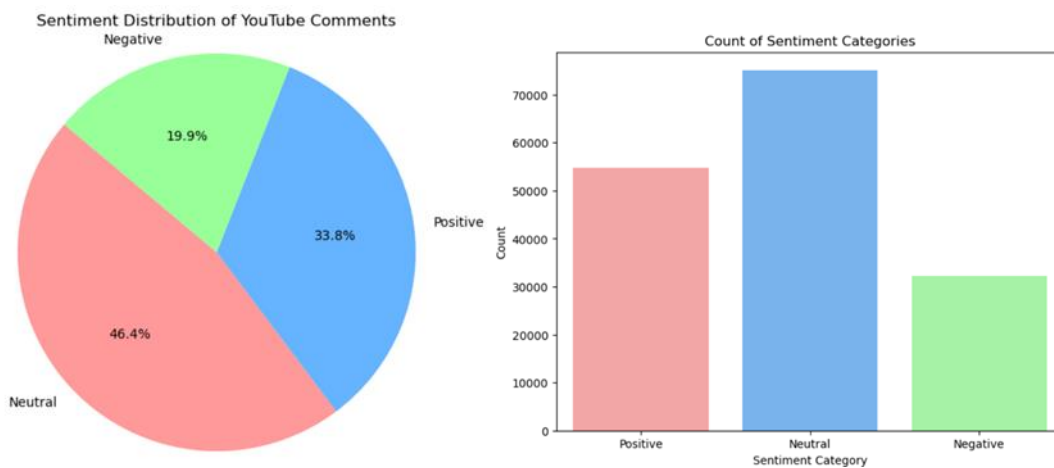


It was observed that about half of the total comments were neutral, 33.8% were positive, and 19.9% were negative.

Sentiment Polarity Distribution
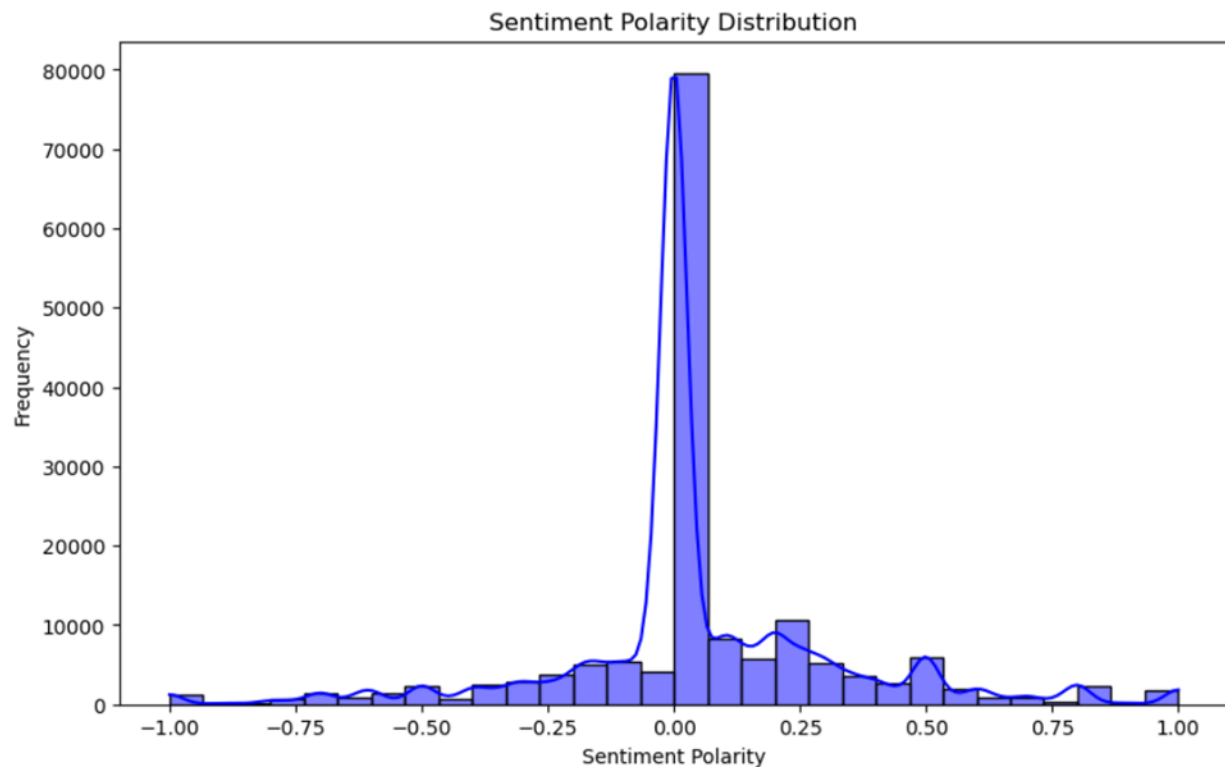
Creating a bar chart of polarity with frequency showed a similar trend. Most of the polarity was at zero meaning comments with neutral sentiment. There were slightly higher peaks on the right side of this graph which were the positive values showing a slight positive sentiment in the comments.

The analysis revealed distinct patterns in the sentiment of YouTube comments related to a specific event. Here are the detailed findings:

The top 5 positive comments reflect significant appreciation and enthusiasm for the event described. Notable examples include phrases like "greatest night in the history of television" and "will smith did wonderful job slapping joker chris." These comments emphasize notable moments and actions, often referring to the event as historic or the performance as commendable. Other comments highlighted the entertainment value with statements such as "hollyweird at its best"

and referenced specific moments with admiration, like "greatest moment of TV history." These comments predominantly use subjective words such as "best," "greatest," "legendary," "awesome," and "perfect," underscoring the high regard in which these viewers held the event.

Conversely, the top 5 negative comments convey strong disapproval and criticism. Examples include "pathetic culture we got here," "jada can't take a joke that smirk and eye rolling is disgusting," and "will smith nasty." These comments often express disdain towards the behavior of the individuals involved and the broader cultural context. The sentiment is further amplified by statements such as "I hope will smith never works again that was disgusting" and criticism towards the consequences faced by the individuals, like "its a shame will smith got banned from Oscar awards." The most subjective words in negative comments include "pathetic," "worst," "terrible," "horrible," and "disgusting," highlighting the strong negative emotions conveyed by these users.

Interestingly, some comments were classified as neutral, yet they often carried implicit negative sentiments. The top 5 neutral comments, such as "pathetic culture we got here" and "jada can't take a joke that smirk and eye rolling is disgusting," overlap with the negative comments. This overlap suggests that even comments categorized as neutral by sentiment analysis might convey underlying negative sentiments. The neutral category did not reveal unique subjective words as distinct as those found in the positive and negative categories.

### Subjectivity in Comments

The analysis also identified the most subjective words used in positive and negative comments. In positive comments, words like "best," "greatest," "legendary," "awesome," and "perfect" were

prevalent, reflecting high praise and admiration. In contrast, negative comments were marked by words like "pathetic," "worst," "terrible," "horrible," and "disgusting," indicating strong disapproval and criticism.



Sentiment Analysis
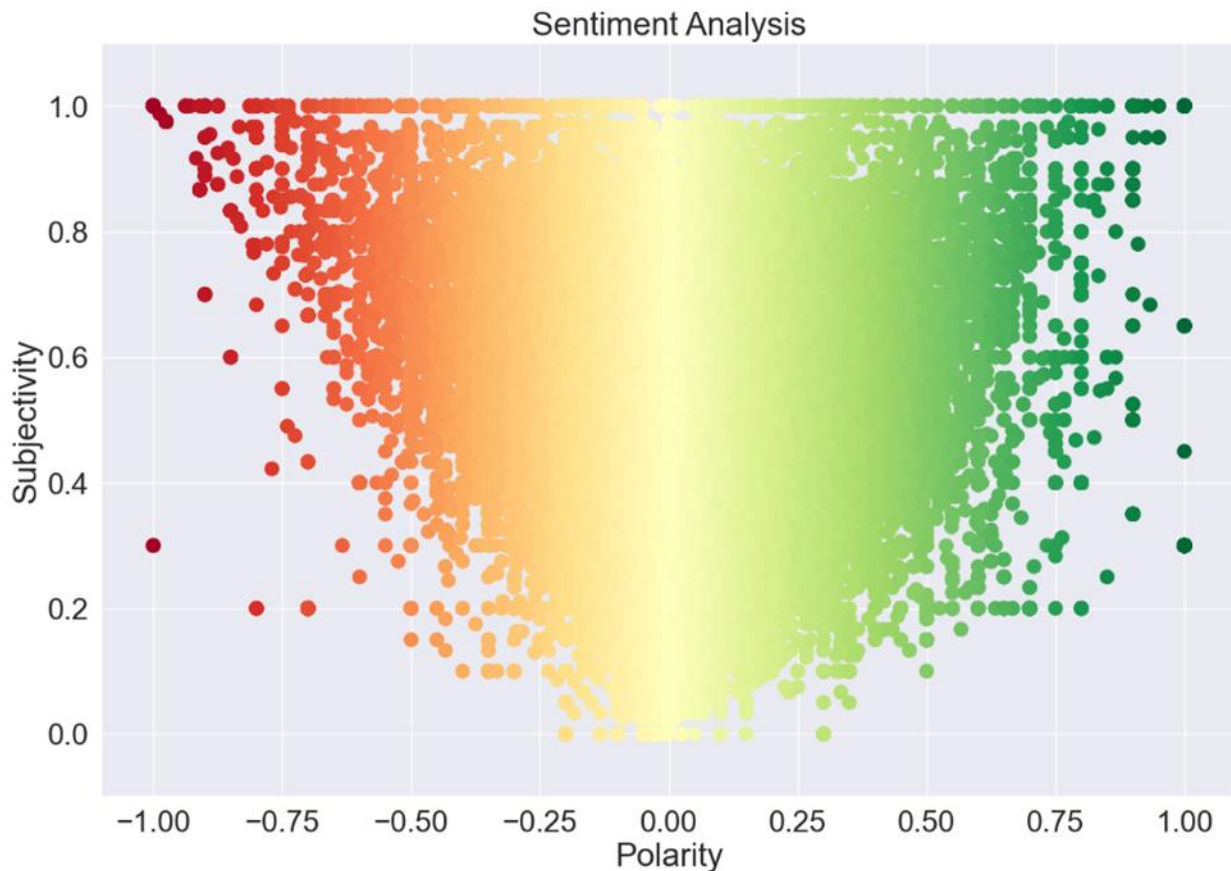
Each point on the plot represents a comment. The position of the point on the x-axis (Polarity) indicates the sentiment of the text (negative, neutral, or positive), and the position on the y-axis (Subjectivity) indicates how subjective or objective the text was. The color gradient from red through yellow to green represents different sentiments or intensities of sentiment, with red indicating negative sentiment, green indicating positive sentiment, and yellow representing neutral or middle-ground sentiment.

The distribution of points suggested a trend where texts became more subjective as they moved towards either extreme of polarity (positive or negative). This could indicate that more emotionally charged comments (either positively or negatively) tended to be more subjective.

Another interesting trend was that the negative polarity grew linearly with the subjectivity in general apart from a few outliers. The positive side showed a non-linear trend, meaning if the polarity started to move towards positive the subjectivity increased as square on average.
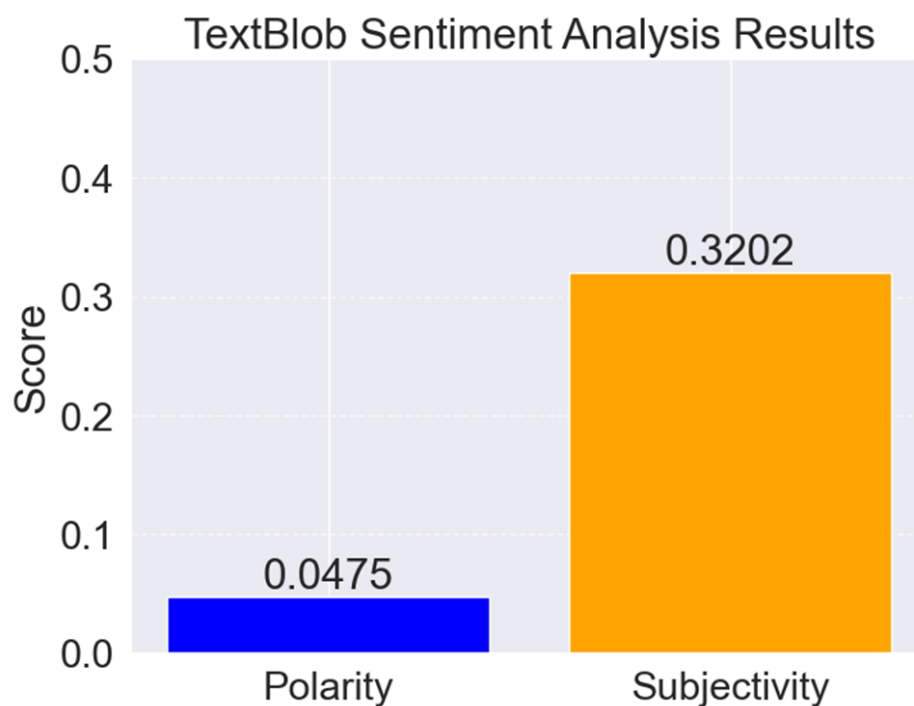


**Overall Sentiment Polarity (TextBlob): 0.0475:**

- The overall sentiment polarity score obtained using TextBlob was 0.0475.

- This value indicated a slightly positive sentiment on average across the analyzed text data.

- A polarity score close to 0 suggested a neutral sentiment, while positive values indicated positive sentiment and negative values indicated negative sentiment.

- Therefore, a score of 0.0475 suggested that, on average, the sentiment expressed in the text data was slightly positive.

**Overall Subjectivity (TextBlob): 0.3202:**

- The overall subjectivity score obtained using TextBlob was 0.3202.

- This value represented the subjectivity of the text, indicating how subjective or opinionated the language was.

- A subjectivity score closer to 0 suggested more objective language, while a score closer to 1 suggested more subjective or opinionated language.

- Therefore, a score of 0.3202 suggested that, on average, the text data contained a moderate level of subjectivity.

The sentiment analysis of YouTube comments revealed a clear dichotomy between positive and negative sentiments, with each category characterized by distinct subjective language. Positive comments were marked by words denoting high praise and admiration, while negative comments were dominated by words expressing strong disapproval and criticism. Neutral comments, despite their classification, often contained implicit negative sentiments. This analysis underscores the importance of examining both sentiment polarity and subjectivity to gain a comprehensive understanding of public opinion in online comments.

# Practical Recommendations

Since that fateful night on which this unfortunate event happened, there have been many popular and unpopular opinions on why, how, and what of the circumstances. There have been just as many discussions on the ideal actions and reactions that should have been taken by both sides. These discussions have been made on the back of analyses of the underlying situations that the people involved found themselves in with layers upon layers of nuances and social factors. However, this paper will provide practical recommendations purely through the analysis of the data examined. The practical recommendations of the paper as follows:

- From the network analysis that was conducted, it appeared that the Violence that occurred at a prestigious and well broadcasted event such as the Oscar's and the extent to which this event was shared on social media platforms had a significant impact on Will Smith's career. Hence, such Violent actions should be avoided as much as possible due to the dire ramifications.
- The sentiment analysis that was conducted on the comments showed that although there was a more positive set of responses as calculated by the sentiment analysis models, a further investigation of the comments flagged as neutral suggests that are more negative reactions by the public towards Will and Jada Smith overall due to their actions. For Jada Smith, her exaggerated reaction to Chris Rock's joke which prompted Will Smith to act so drastically saw some criticism. Hence, another

recommendation that can be gathered from the data is that provocative facial

reactions in public should be avoided.

- Additionally, the wordcloud analysis of the dataset found there to be significant

  mentions of Will Smith's verbal volleys at Chris Rock after the physical altercation.

  This implies that the public also found the verbal reactions of Will Smith to be

  emphatic at the very least. It can be argued that the data shows such language

  draws negative attention as well and should be avoided.

# Limitations of the Study

Although word cloud analysis is a popular tool for visualizing the frequency of words within a text,

it has some limitations in the context of analyzing complex datasets. It does not provide

information about the relationships or the context between words. Moreover, it would display

words in isolation, stripping them of their context. For instance, words like "wife" or "joke" or

"got" might appear very frequent but without indicating the sentiment or the specific discussions

surrounding these terms. In another hand, significant but less frequent terms may be completely

overlooked.

Sentiment analysis and LDA analysis might give out more accurate and unbiased results. It

typically reduces complex emotions to simple positive, negative, or neutral categories, which can

oversimplify the true sentiments. Determining the optimal number of topics in LDA is often

subjective. This method can struggle with words that have multiple meanings leading to

misclassifying thematic content.

The structure of the dendrogram is sensitive to the method of distance calculation and linkage used in hierarchical clustering, which could affect the interpretation of how topics are related. The distances in the dendrogram provide a visual indication of similarity but may not capture all nuances, especially with abstract or complex topics like emotions or sentiments. Interpretation remains somewhat subjective, especially in deciding how closely topics are related based on the clustering outputs which usually highlighted the most frequencies of words in public comments.

# References

Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8*(4), e1253. https://doi.org/10.1002/widm.1253

Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press. https://doi.org/10.1017/CBO9781139084789

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems, 28*(2), 15-21. https://doi.org/10.1109/MIS.2013.30

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley. https://doi.org/10.1002/9780470316801

Heimerl, F., Lohmann, S., Lange, S., & Ertl, T. (2014). Word cloud explorer: Text analytics based on word clouds. *Proceedings of the 47th Hawaii International Conference on System Sciences*, 1833-1842. https://doi.org/10.1109/HICSS.2014.231