

Tuesday, 15 November

School of Advanced Sciences, VIT, Vellore

SET PROJECT PRESENTATION:

A Sentiment Analysis Case Study to Understand How A YouTuber Can Derive Decision Insights From Comments

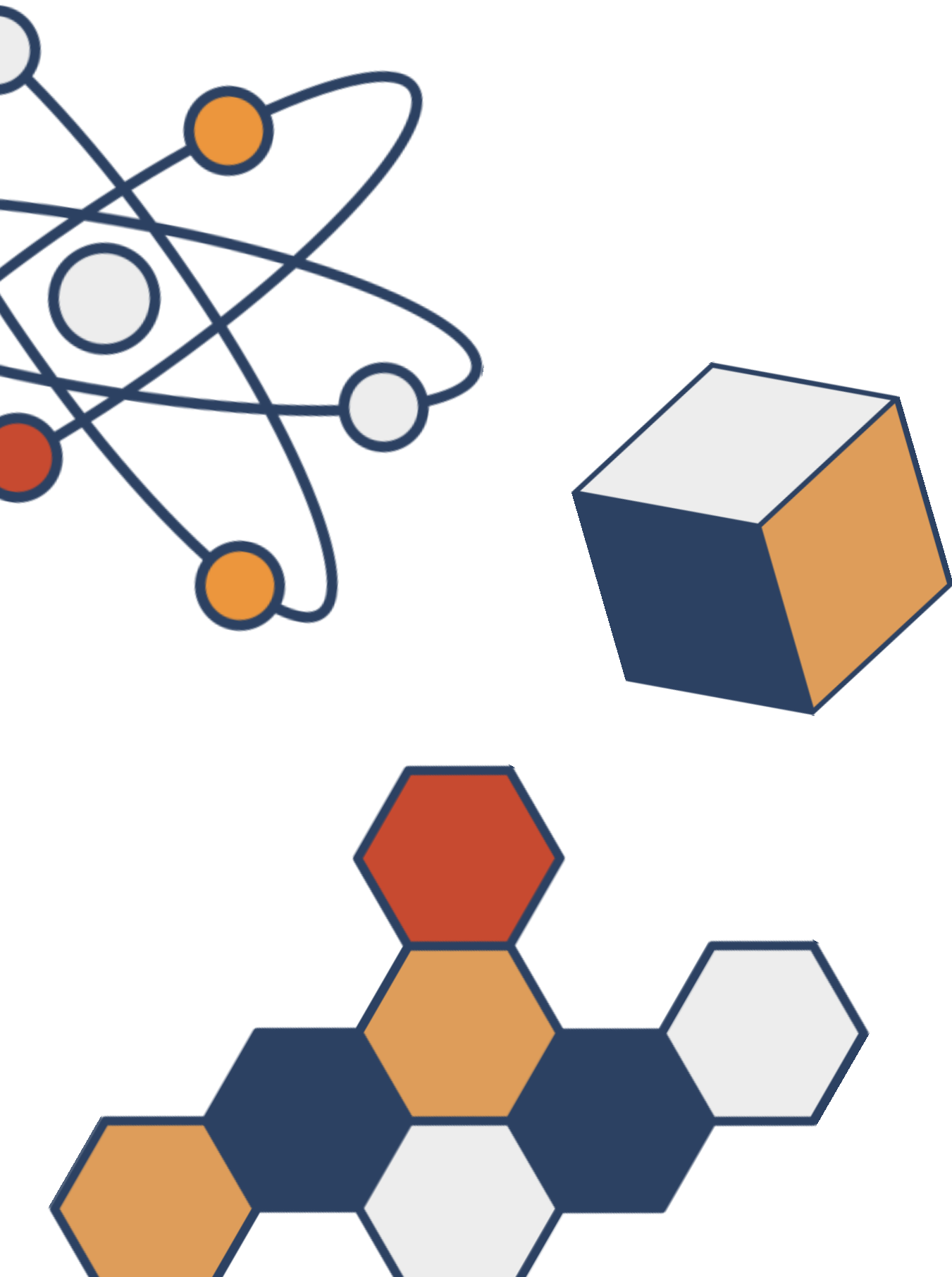
STUDENTS

Chandreyi : 19MIY0031
Pavitra : 19MIY0035

ADVISOR

SRV Prasad Bhuvanagiri





- As the topic suggests, our research is based on the comments on YouTube videos
- YouTube is considered to be one of the largest social media platform
- Millions of video based content are uploaded eachday having billions of comments
- A YouTuber aims to give his/her viewers the best content
- Comments are a way to understand what the viewers want

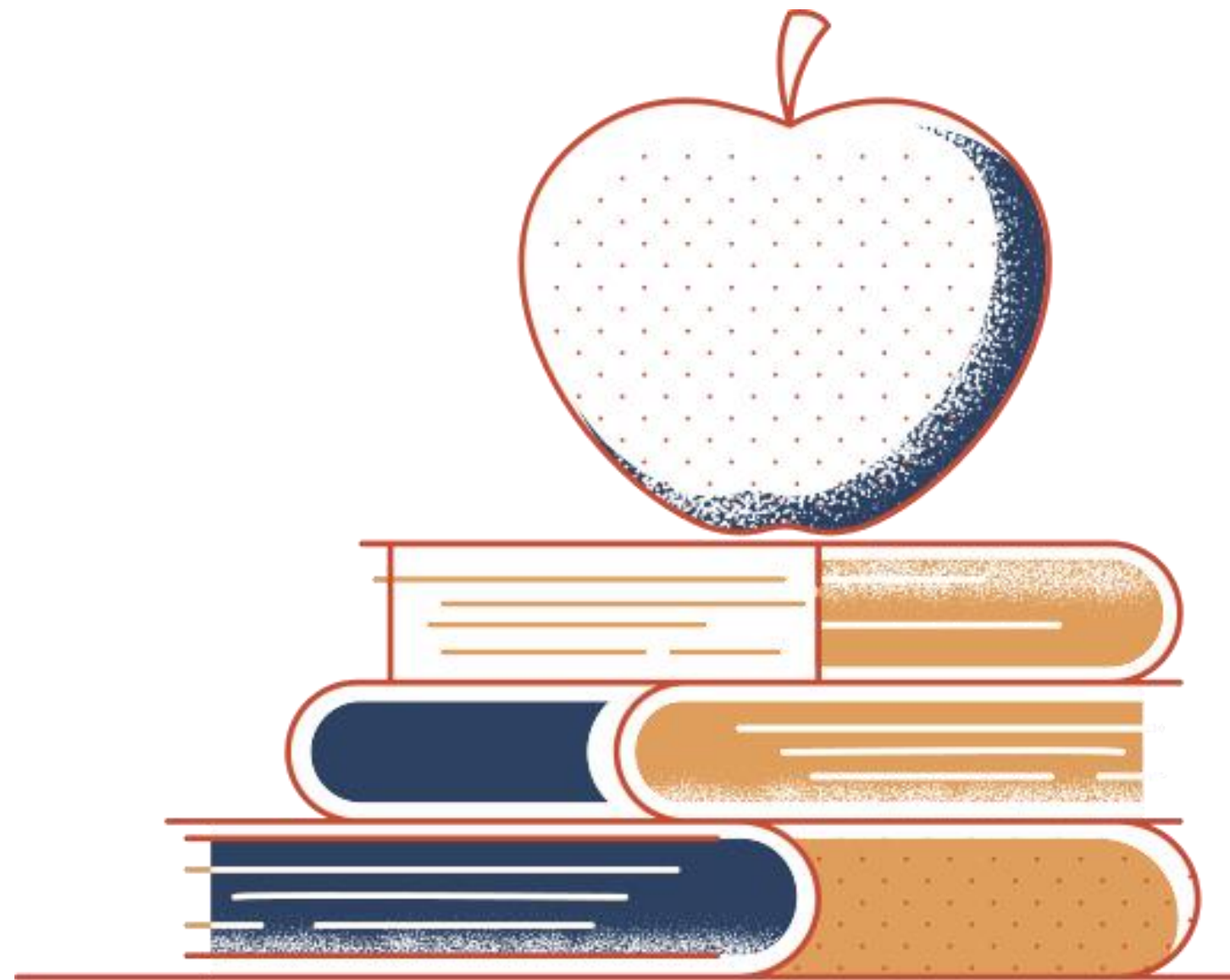
Table of Contents		
		Page
I	Abstract	2
II	Introduction	4
III	Research Background and Motivation	6
IV	Methodology	8
V	Results and Conclusion	15

- A simple way to analyse the reputation of a video is by the number of likes and dislikes
- If the number of likes is much higher than the number of dislikes, it is good content
- but a high number of dislikes compared to likes usually means bad content
- However, there are a few ways to measure reputation quantitatively. This raises the importance of automatically extracting sentiments and opinions expressed in social media
- Therefore, sentiment analysis on the data extracted from the comments section of a video, can be used to examine YouTube users' perceptions of video content.

- Going through each and every comment can be a bit over whelming since on an average an youtube video can get upto 1000 comments
- Therefor our research model aims at providing the YouTuber a way to get an understanding of the viewers choices through the visual representation of the sentiment analysis performed on the comments extracted.



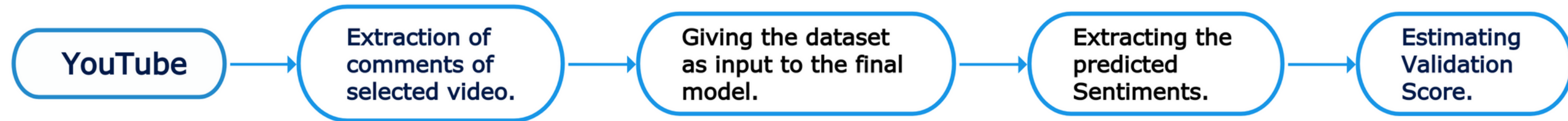
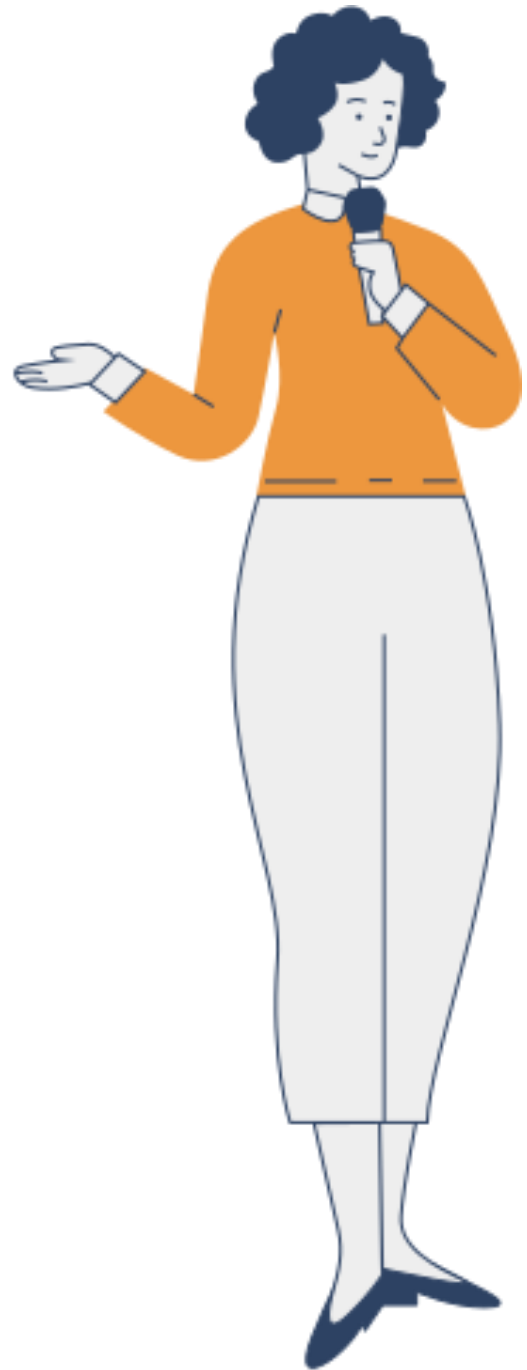
- The existing research papers on sentiment analysis focused on the data extraction and performing sentiment analysis on the data extracted.
- To the best of our knowledge, a quick, effective and visually pleasing method mainly for the YouTubers/ Influencers to be able to detect information and take future decisions based on the comments, which can help them to get the sentiment distribution of that particular video doesn't exist.





Tools used:-


- Machine Learning
- Scikit-learn library
- Selenium, BeautifulSoup
- Natural Language Toolkit.



Data extraction

Beautiful Soup makes it simple to gather data from websites, sits atop an HTML or XML parser, providing Python-base idioms for iterating, searching, and modifying the parse tree.





 Pinned by Nutshell

Nutshell 1 year ago

What other finance-related videos should we make? 🌱 🌱 What'd you like us to break down next? 📌


👍 127 💬 Reply

▼  • 75 replies

 **santhosh peddi** 11 months ago


Great way of making new generation understand history with stories and facts .
Amazing work 😊


👍 3 💬 Reply

 **Himanshu Kumar** 1 year ago

Randomly landed on this channel while watching another video related to politics. Very good content sir, I liked it very much.


👍 7 💬 ❤️ Reply

▼  • 3 replies

 **yashsavplayz** 1 year ago


I am still counfused why this 2 k people will dislike this masterpiece

👍 1 💬 Reply

 **Prateek Verma** 10 months ago


This is something amazing I've saw on YouTube these days ❤️ ... Loved it


👍 1 💬 Reply

 **Uswa Rizvi** 1 year ago

Love this channel. Just amazing 🤔 ❤️

👍 4 💬 ❤️ Reply

▼  • 1 reply

 **Gauravi Chandak** 1 year ago

Yes please! Economics is so much fun but a bit hard to get and you make it a tad bit easier.

👍 💬 Reply

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1		Comment																	
2	0	What other finance-related videos should we make? What'd you like us to break down next?																	
3	1	Great																	
4	2	This is something amazing I've saw on YouTube these days ... Loved it																	
5	3	Randomly landed on this channel while watching another video related to politics. Very good content sir, I liked it very much.																	
6	4	I am still counfused why this 2 k people will dislike this masterpiece																	
7	5	Yes please! Economics is so much fun but a bit hard to get and you make it a tad bit easier.																	
8	6	Love this channel. Just amazing ĩ,																	
9	7	Loved this video and I love topics of finance and accounting or business related. So please make more videos on this topics.ĩ,																	
10	8	Chota Packet, Bada Dhamaka ... Real good job you guys																	
11	9	Really informative.. Would like to watch more of this kind..																	
12	10	Thank you so much for providing information to us																	
13	11	Yes you should make more of such finance related videos																	
14	12	Very good and clear accent. No too much drama, just to the point.																	
15	13	Can you make a video on why and how DEMONETIZATION failed in India??																	
16	14	we need such videos!!!!!!																	
17	15	You are such best channels on YouTube because of unique concept and ideas keep doing this																	
18	16	I thought black money is just money not printed by RBI until I watch this video. Thanks man for a great piece of information.																	
19	17	Oh buoy ,, you people are literally so underrated. Amazing videos . Thanks for such a factful n useful video.																	
20	18	Can you make a video about how many people pay tax in India and how much of total transactions are in black and most popular cases where people got arrested.																	
21	19	We really like this kind of stuff																	
22	20	Nice video. Keep it up																	
23	21	Hey.. These videos are just amazing... Keep going... I would suggest you to create content that could be of great help to the common man... I mean to say in India people get into major troubles jus																	
24	22	The video is really amazing.																	
25	23	Great initiative want more videos like this																	
26	24	Absolutely a great initiative by Nutshell.																	
27	25	Thanks Appreciated Keep it up																	

Text Preprocessing

The target sentiments were decided to be Positive, Negative and Neutral, and the dataset was transformed.

Text sentiment analysis in our model is carried out using the VADER (Valence Aware Dictionary for Sentiment Reasoning) model, which is sensitive to both the polarity (positive/negative) and intensity (strong) of emotion



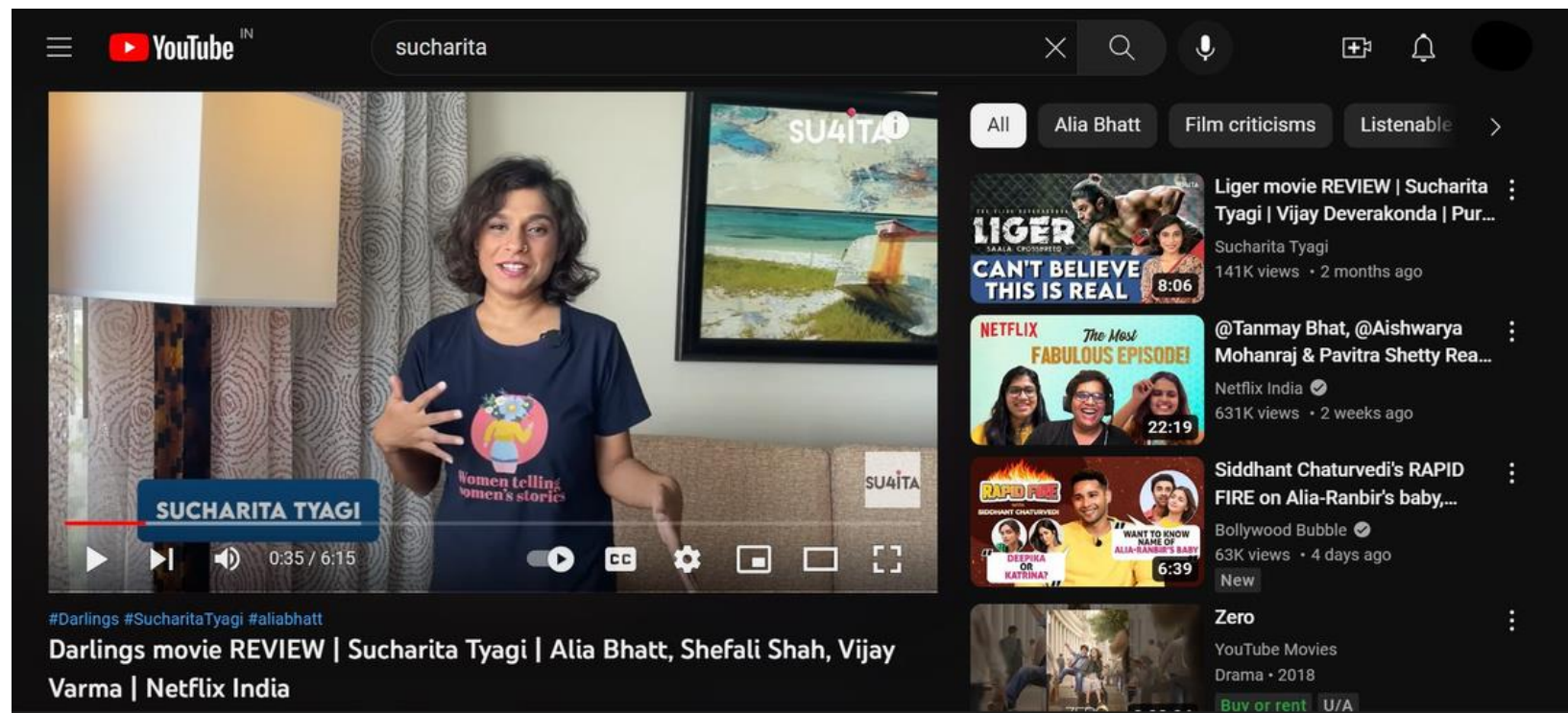
Validation

- Our model used accuracy score, and root mean squared error as our accuracy metrics.
- The basic intuitive notion of the formula for accuracy score is:



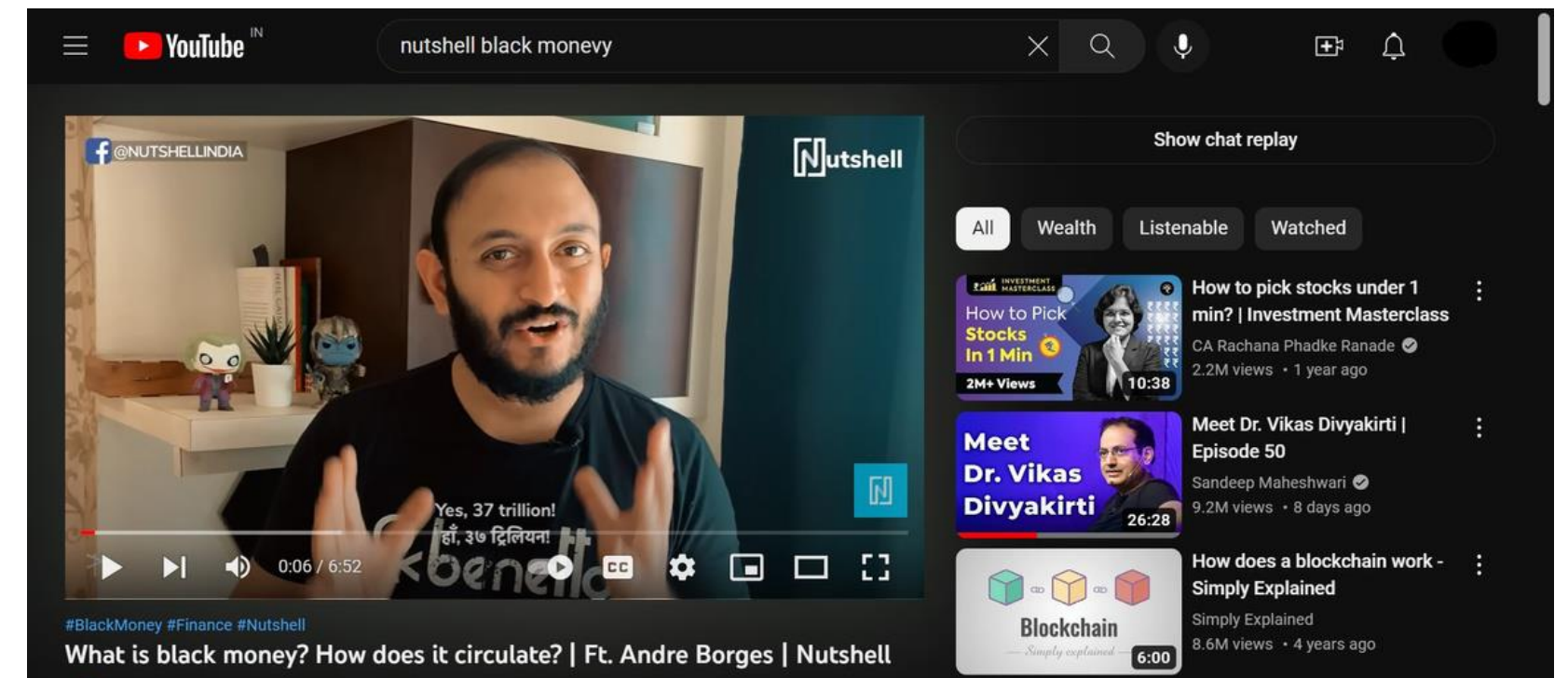
```
In [32]: ▶ def score(model, title = "Default"):
            model.fit(X_train, y_train)
            preds = model.predict(X_test)
            # print(confusion_matrix(y_test, preds))
            rmse = round(mean_squared_error(y_test, y_pred, squared=False), 5)
            acc = round(accuracy_score(y_test, y_pred), 5)
            print('RMSE for', title, ':', rmse, '\n')
            print('Accuracy score for', title, ':', acc, '\n')
```


Movie Reviews by Sucharita -> Darlings (the 5th of August, 2022)



<i>Account</i>	Sucharita Tyagi
<i>Subscribers</i>	62.5K
<i>Video Caption</i>	Darlings movie REVIEW Sucharita Tyagi Alia Bhatt, Shefali Shah, Vijay Varma Netflix India
<i>Published Date</i>	The 5th of August, 2022
<i>Video URL</i>	https://www.youtube.com/watch?v=IPvXZz7m9sI&t=2s
<i>Total Comments</i>	193 (as of the 2nd of November 2022)
<i>Scrapped Comments</i>	119

Nutshell -> Black Money (the 21st of May, 2021)



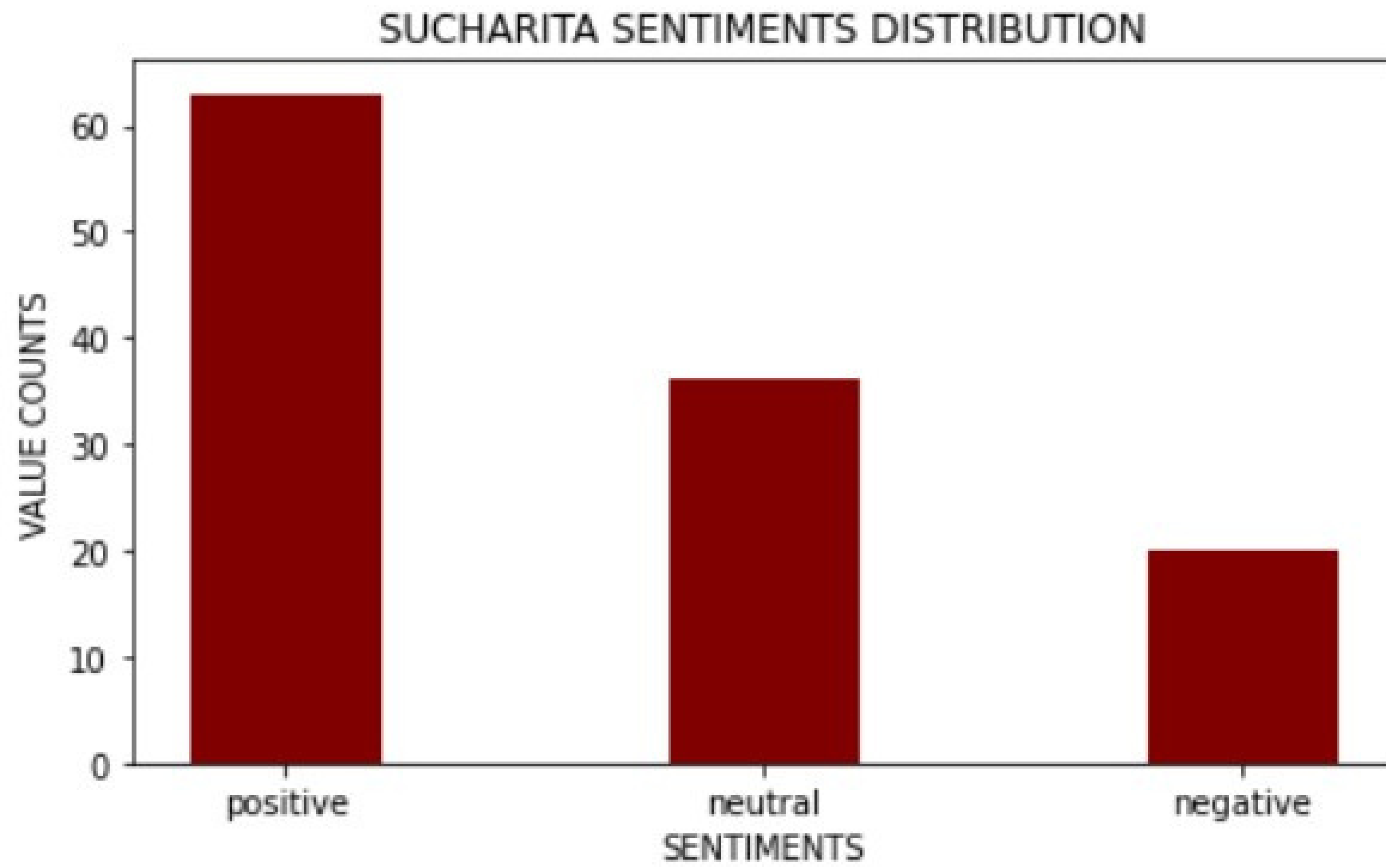
<i>Account</i>	Nutshell
<i>Subscribers</i>	243K
<i>Video Caption</i>	What is black money? How does it circulate? Ft. Andre Borges Nutshell
<i>Published Date</i>	The 21st of May, 2021
<i>Video URL</i>	https://www.youtube.com/watch?v=uEawmeO2gOY&t=7s
<i>Total Comments</i>	228 (as of the 2nd of November 2022)
<i>Scrapped Comments</i>	93

Case Study 1

Proceeding towards the model description and analysis of this dataset, we found that the processed data had 20 negative classes, 63 positive classes and 36 neutral classes.

Finally, on this dataset, applying several machine learning classifiers by similarly training the models with 70 per cent of the data and saving the rest 30 per cent for testing, we get the best accuracy score of 0.9781 from the Gaussian Naïve Bayes Classifier. Hence, we can also conclude that the distribution we initially got is almost correct. So, it means that approximately 48 per cent of its viewers have a favorable opinion, 37 per cent have a neutral opinion and the rest 15 per cent have a negative opinion about the video. Figure 12 shows the same diagrammatically.



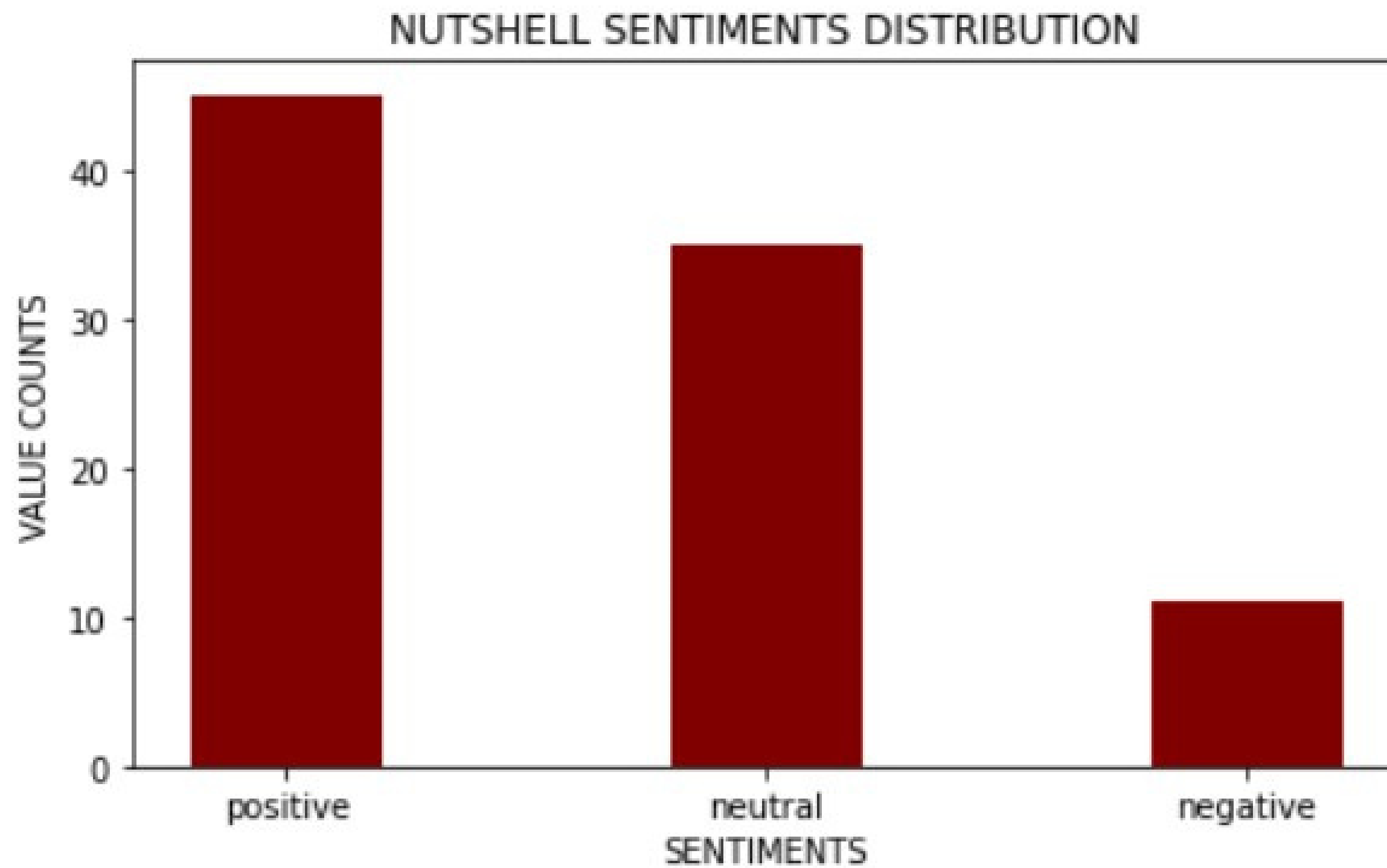


Case Study 2

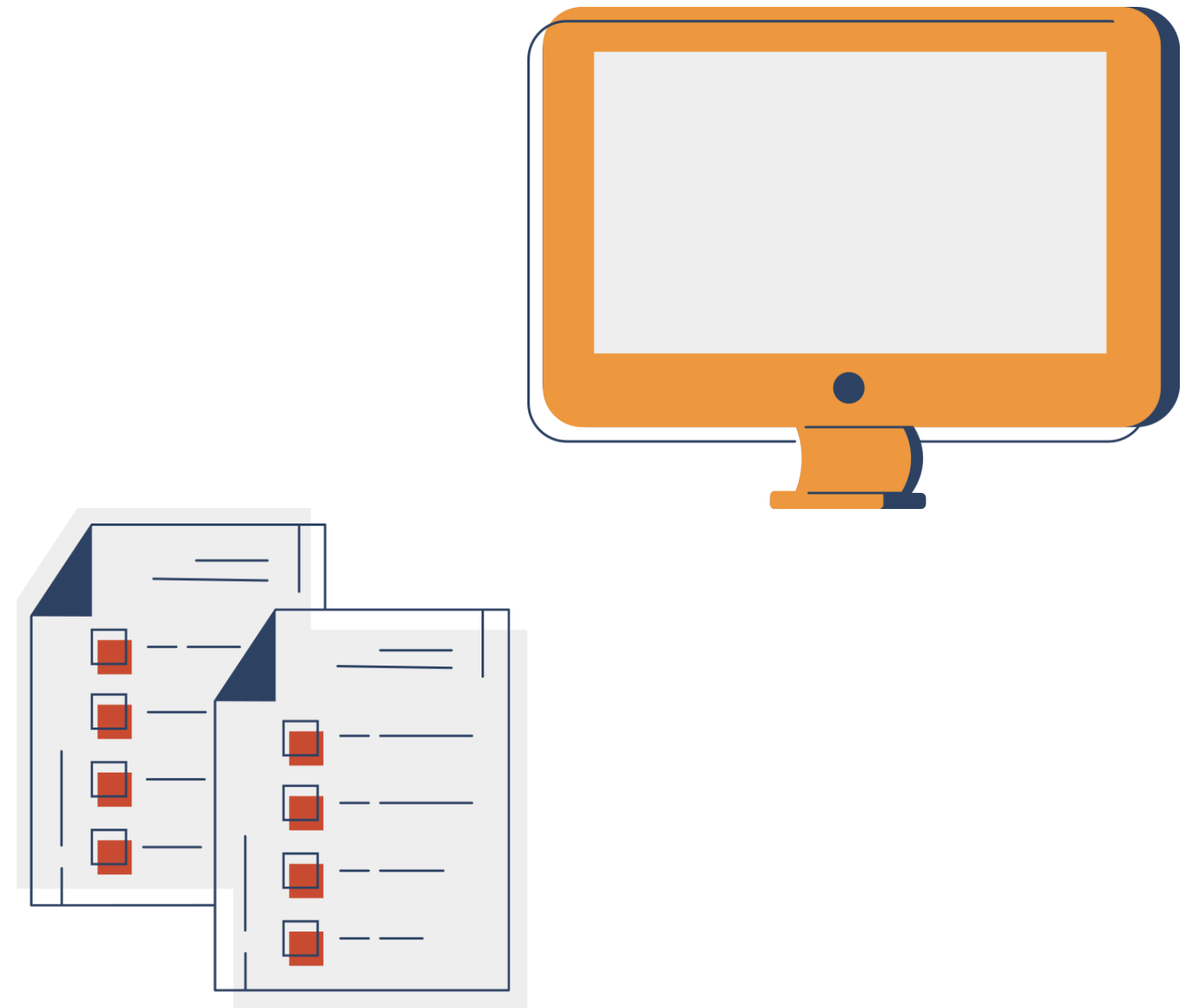
For this video, proceeding towards the model description and analysis, we found that the processed data had 11 negative classes, 45 positive classes and 35 neutral classes.

Finally, on this dataset, applying several machine learning classifiers by similarly training the models with 70 per cent of the data and saving the rest 30 per cent for testing, we get the best accuracy score of 0.9781 from the Gaussian Naïve Bayes Classifier. Hence, we can also conclude that the distribution we initially got is almost correct. So, it means that approximately 48 per cent of its viewers have a favorable opinion, 37 per cent have a neutral opinion and the rest 15 per cent have a negative opinion about the video.





After performing these two case studies, we can say that most of the comments were on the positive side of both videos. Moreover, the respective viewers liked the contents of the two videos and want more of the same kind in future. The ratio of negative comments was meagre, i.e., less than 20 per cent for both videos, which is an ideal scenario for the content creators. Furthermore, when they want suggestions for other content creation, they can go through the neutral comments, which generally contain suggestions from the viewers.



The End

Thank you

