```
In [29]:   !pip install youtube_transcript_api
           from youtube_transcript_api import YouTubeTranscriptApi

           Requirement already satisfied: youtube_transcript_api in /usr/local/lib/python3.10/site-
           packages (0.4.4)
           Requirement already satisfied: requests in /usr/local/lib/python3.10/site-packages (from
           youtube_transcript_api) (2.28.1)
           Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/site-packages
           (from requests->youtube_transcript_api) (3.4)
           Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/site-pack
           ages (from requests->youtube_transcript_api) (2022.9.14)
           Requirement already satisfied: charset-normalizer<3,>=2 in /usr/local/lib/python3.10/sit
           e-packages (from requests->youtube_transcript_api) (2.1.1)
           Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.10/site-p
           ackages (from requests->youtube_transcript_api) (1.26.12)

           [notice] A new release of pip available: 22.2.2 -> 22.3
           [notice] To update, run: python3.10 -m pip install --upgrade pip
```

```
In [88]:   # Import all necessary Libraries
           import re
           import spacy
           import nltk
           import numpy as np
           from nltk.stem import WordNetLemmatizer
           from nltk.tokenize import word_tokenize, sent_tokenize
           from keras.utils import to_categorical
           from sklearn.model_selection import train_test_split
           from keras.models import Sequential
           from keras.layers import LSTM, Dense, GRU, Embedding
           from keras.callbacks import EarlyStopping, ModelCheckpoint
           from keras_preprocessing.sequence import pad_sequences
           from textblob.en import Spelling
           from textblob import TextBlob
           from nltk import bigrams, trigrams
           from collections import Counter, defaultdict
           from sklearn.feature_extraction.text import CountVectorizer
           from sklearn.naive_bayes import MultinomialNB
           from sklearn.pipeline import Pipeline
```

```
In [31]:   # Access youtube video from video id from URL
           youtube_link = "https://www.youtube.com/watch?v=vsZYupeZ9Ao&ab_channel=ALTEXTV"
           # youtube_link = input('Enter youtube link:')
           video_id = re.findall(r'v=[a-zA-Z0-9]+', youtube_link)[0][2:]
           print(youtube_link, video_id)

           https://www.youtube.com/watch?v=vsZYupeZ9Ao&ab_channel=ALTEXTV vsZYupeZ9Ao
```

```
In [32]:   # Get the transcript from youtube video
           youtube_api = YouTubeTranscriptApi.get_transcript(video_id)
           print(youtube_api)
```

[{'text': 'our scene alive is a new action movie', 'start': 0.08, 'duration': 4.32}, {'text': 'that is coming to the big screen for you', 'start': 2.48, 'duration': 3.68}, {'text': 'all to go and check out this movie stars', 'start': 4.4, 'duration': 3.84}, {'text': 'gerard butler and if you guys know his', 'start': 6.16, 'duration': 3.6}, {'text': "movie you know pretty much what you're", 'start': 8.24, 'duration': 5.2}, {'text': 'gonna get this movie we follow him as he', 'start': 9.76, 'duration': 5.919}, {'text': 'is looking for his wife who has been', 'start': 13.44, 'duration': 4.08}, {'text': 'kidnapped', 'start': 15.679, 'duration': 3.121}, {'text': 'and', 'start': 17.52, 'duration': 3.519}, {'text': "he's gone to do it by himself it's red", 'start': 18.8, 'duration': 3.68}, {'text': "bench phil so i'm going to tell you guys", 'start': 21.039, 'duration': 2.641}, {'text': 'my thoughts on this movie let me know in', 'start': 22.48, 'duration': 2.32}, {'text': 'the comments below what do you guys', 'start': 23.68, 'duration': 2.88}, {'text': "think of this movie or if you haven't", 'start': 24.8, 'duration': 3.2}, {'text': 'seen it what are your thoughts on the', 'start': 26.56, 'duration': 3.119}, {'text': 'trailer are you excited smash it down', 'start': 28.0, 'duration': 2.96}, {'text': 'below smash that like button click', 'start': 29.679, 'duration': 2.961}, {'text': "subscribe and let's jump into this", 'start': 30.96, 'duration': 6.439}, {'text': 'review of last scene alive', 'start': 32.64, 'duration': 4.759}, {'text': 'okay so last scene alive is a revenge', 'start': 45.52, 'duration': 5.44}, {'text': "field that's to its core", 'start': 48.48, 'duration': 4.8}, {'text': "that's what this film is it's a complete", 'start': 50.96, 'duration': 5.279}, {'text': 'and utter revenge film so', 'start': 53.28, 'duration': 3.84}, {'text': 'me', 'start': 56.239, 'duration': 2.721}, {'text': "i'm a sucker for those revenge fields so", 'start': 57.12, 'duration': 3.919}, {'text': "i mean we got john wick john wick's a", 'start': 58.96, 'duration': 4.88}, {'text': "classic revenge film we've got rambo", 'start': 61.039, 'duration': 6.08}, {'text': "revenge films we've got all these other", 'start': 63.84, 'duration': 5.44}, {'text': "revenge films and i'm pretty sure most", 'start': 67.119, 'duration': 5.36}, {'text': "people enjoy them because they're just", 'start': 69.28, 'duration': 4.879}, {'text': 'so much fun and so much entertainment', 'start': 72.479, 'duration': 3.68}, {'text': 'and we like to see', 'start': 74.159, 'duration': 3.121}, {'text': 'the', 'start': 76.159, 'duration': 2.401}, {'text': 'good guys', 'start': 77.28, 'duration': 3.6}, {'text': "absolutely murder bad guys it's just", 'start': 78.56, 'duration': 4.239}, {'text': "it's in our dna it's in our nature", 'start': 80.88, 'duration': 3.84}, {'text': "that's what we want to see", 'start': 82.799, 'duration': 4.081}, {'text': 'when you put in jared butler', 'start': 84.72, 'duration': 3.759}, {'text': "you know what you're going to get i mean", 'start': 86.88, 'duration': 4.4}, {'text': 'angel has fallen the whole has fallen', 'start': 88.479, 'duration': 4.241}, {'text': "trilogy you know what you're going to", 'start': 91.28, 'duration': 4.4}, {'text': 'get from this guy he is the modern day', 'start': 92.72, 'duration': 5.039}, {'text': "liam neeson that's what he is he does", 'start': 95.68, 'duration': 3.92}, {'text': 'these little action films like this', 'start': 97.759, 'duration': 4.241}, {'text': 'these lower budget ones', 'start': 99.6, 'duration': 4.559}, {'text': 'just like liam neeson used to', 'start': 102.0, 'duration': 3.759}, {'text': "that's what i put these in that in that", 'start': 104.159, 'duration': 3.6}, {'text': "sort of category this one's no different", 'start': 105.759, 'duration': 5.761}, {'text': "it's fun it is fun and i tell you what", 'start': 107.759, 'duration': 6.081}, {'text': 'it has a lot more heart than i thought', 'start': 111.52, 'duration': 4.239}, {'text': 'this movie was going to have', 'start': 113.84, 'duration': 3.52}, {'text': 'way more heart', 'start': 115.759, 'duration': 3.04}, {'text': 'it started to pull up my strengths', 'start': 117.36, 'duration': 2.719}, {'text': 'because', 'start': 118.799, 'duration': 3.6}, {'text': "jared butler's character", 'start': 120.079, 'duration': 4.72}, {'text': 'i actually felt for him they did a', 'start': 122.399, 'duration': 3.921}, {'text': 'fantastic job', 'start': 124.799, 'duration': 4.481}, {'text': 'humanizing him to the point where', 'start': 126.32, 'duration': 5.44}, {'text': "we don't just see him as this", 'start': 129.28, 'duration': 4.64}, {'text': "terrible husband we don't see him as", 'start': 131.76, 'duration': 4.559}, {'text': "this terrible human being we don't he's", 'start': 133.92, 'duration': 5.6}, {'text': 'not even especially skilled', 'start': 136.319, 'duration': 4.961}, {'text': "or anything like that he's just a", 'start': 139.52, 'duration': 3.28}, {'text': 'regular', 'start': 141.28, 'duration': 3.36}, {'text': 'real estate agent', 'start': 142.8, 'duration': 3.92}, {'text': "guy that's what he is he does real", 'start': 144.64, 'duration': 3.84}, {'text': "estate that's all it is", 'start': 146.72, 'duration': 2.8}, {'text': 'but', 'start': 148.48, 'duration': 3.039}, {'text': 'what are they they did so well is they', 'start': 149.52, 'duration': 4.6

4}, {'text': 'showed us the whole background as we', 'start': 151.519, 'duration': 4.7
2}, {'text': 'move forward we see the whole background', 'start': 154.16, 'duration': 3.
84}, {'text': "of their relationship and that's what", 'start': 156.239, 'duration': 3.3
6}, {'text': "it's really built on", 'start': 158.0, 'duration': 3.36}, {'text': 'we see
the ups and downs', 'start': 159.599, 'duration': 2.881}, {'text': 'and we see', 'star
t': 161.36, 'duration': 3.36}, {'text': 'this guy that just really cares about', 'star
t': 162.48, 'duration': 4.64}, {'text': 'his wife and when you take that away', 'start':
164.72, 'duration': 4.879}, {'text': 'from him he is willing to do anything', 'start': 1
67.12, 'duration': 4.88}, {'text': 'and everything to get it back', 'start': 169.599, 'd
uration': 3.841}, {'text': "and that's what i like about it they", 'start': 172.0, 'dura
tion': 3.44}, {'text': "don't go over the top with him doing", 'start': 173.44, 'duratio
n': 4.879}, {'text': 'these insane skills of', 'start': 175.44, 'duration': 6.159}, {'te
xt': 'like a secret agent a secret', 'start': 178.319, 'duration': 5.441}, {'text': "sol
dier anything like that he doesn't do", 'start': 181.599, 'duration': 4.72}, {'text': 't
hat he is just a generic guy and he', 'start': 183.76, 'duration': 5.36}, {'text': 'does
what a generic person would do', 'start': 186.319, 'duration': 5.92}, {'text': "so it's
not polished and i respect that", 'start': 189.12, 'duration': 5.28}, {'text': 'because
they could have easily', 'start': 192.239, 'duration': 4.961}, {'text': 'wrote it to whe
re he had served or done', 'start': 194.4, 'duration': 4.72}, {'text': "something and h
e's like a special agent", 'start': 197.2, 'duration': 3.52}, {'text': "and that's why h
e can do all these cool", 'start': 199.12, 'duration': 3.52}, {'text': "skills or anythi
ng like but it's not", 'start': 200.72, 'duration': 4.0}, {'text': "that's not what it i
s and they really", 'start': 202.64, 'duration': 5.2}, {'text': 'made it evident that yo
u could actually', 'start': 204.72, 'duration': 5.2}, {'text': 'follow along and underst
and his', 'start': 207.84, 'duration': 4.24}, {'text': 'motivation to why he was doing i
t and', 'start': 209.92, 'duration': 4.72}, {'text': 'you could understand', 'start': 21
2.08, 'duration': 4.64}, {'text': "exactly where he's coming from and when", 'start': 21
4.64, 'duration': 4.239}, {'text': 'they showed us', 'start': 216.72, 'duration': 3.84},
{'text': 'all the stuff that happened with him and', 'start': 218.879, 'duration': 3.20
1}, {'text': 'his character', 'start': 220.56, 'duration': 4.16}, {'text': "and how brok
en he is and how he's trying", 'start': 222.08, 'duration': 4.0}, {'text': 'to build it
up and he just wants to be', 'start': 224.72, 'duration': 3.68}, {'text': "this great hu
sband he's willing to do", 'start': 226.08, 'duration': 3.92}, {'text': "like some of th
e stuff he's gone through", 'start': 228.4, 'duration': 3.68}, {'text': "he's willing to
still do", 'start': 230.0, 'duration': 2.799}, {'text': 'like', 'start': 232.08, 'durati
on': 3.28}, {'text': "be there for her like it's insane", 'start': 232.799, 'duration':
4.321}, {'text': 'that really gets you on his side so', 'start': 235.36, 'duration': 4.2
39}, {'text': 'whenever he goes around and does kill', 'start': 237.12, 'duration': 4.5
6}, {'text': 'anyone anything like that', 'start': 239.599, 'duration': 4.56}, {'text':
'you just push it to the side and go yeah', 'start': 241.68, 'duration': 5.44}, {'text':
'justified justified so', 'start': 244.159, 'duration': 4.16}, {'text': 'that really rea
lly worked jamie', 'start': 247.12, 'duration': 3.839}, {'text': 'alexander as the wife
she was good for', 'start': 248.319, 'duration': 4.321}, {'text': 'the what the whole 10
minutes she was on', 'start': 250.959, 'duration': 4.48}, {'text': 'screen for so she wa
s great the cop i', 'start': 252.64, 'duration': 4.4}, {'text': 'enjoyed him as well',
'start': 255.439, 'duration': 3.681}, {'text': 'what they really did is they did the co
p', 'start': 257.04, 'duration': 4.159}, {'text': 'storyline and his storyline as well y
ou', 'start': 259.12, 'duration': 3.6}, {'text': 'know how they call the cops first the
n', 'start': 261.199, 'duration': 3.121}, {'text': 'the cops take forever so he goes [\x
a0__\xa0] it', 'start': 262.72, 'duration': 3.6}, {'text': "i'm gonna go do it myself th
at happens", 'start': 264.32, 'duration': 3.92}, {'text': 'in this as well but i like th
e cop as', 'start': 266.32, 'duration': 4.0}, {'text': 'well he does a good job', 'star
t': 268.24, 'duration': 4.48}, {'text': 'the cinematography is great i really', 'start':
270.32, 'duration': 4.319}, {'text': 'enjoy the cinematography gorgeous', 'start': 272.7
2, 'duration': 3.36}, {'text': 'locations', 'start': 274.639, 'duration': 4.641}, {'tex
t': 'the camera work was outstanding as well', 'start': 276.08, 'duration': 5.679}, {'te
xt': 'this visual effects there is a massive', 'start': 279.28, 'duration': 4.88}, {'tex
t': 'explosion and you can tell that it is', 'start': 281.759, 'duration': 3.281}, {'tex
t': 'quite', 'start': 284.16, 'duration': 3.44}, {'text': 'evidently low budget which is
fine', 'start': 285.04, 'duration': 3.52}, {'text': 'because', 'start': 287.6, 'duratio
n': 2.879}, {'text': 'these movies are low budget so you kind', 'start': 288.56, 'durati
on': 3.919}, {'text': "of got to expect that the effects aren't", 'start': 290.479, 'dur

ation': 4.561}, {'text': 'going to be like', 'start': 292.479, 'duration': 4.561}, {'text': 'comic book movie level', 'start': 295.04, 'duration': 4.48}, {'text': "these don't have budgets of 50 plus", 'start': 297.04, 'duration': 3.76}, {'text': 'million', 'start': 299.52, 'duration': 3.36}, {'text': 'these are the low low budgets and i like', 'start': 300.8, 'duration': 3.6}, {'text': "these low budget action fields they're", 'start': 302.88, 'duration': 3.599}, {'text': 'cool we need movies like this this is', 'start': 304.4, 'duration': 4.72}, {'text': 'just a nice fun time at the theater', 'start': 306.479, 'duration': 4.56}, {'text': 'and it has some suspense to it as well', 'start': 309.12, 'duration': 3.68}, {'text': "there's a there's a scene", 'start': 311.039, 'duration': 3.121}, {'text': 'where they do it really really well', 'start': 312.8, 'duration': 2.88}, {'text': "where he's actually going around the", 'start': 314.16, 'duration': 4.16}, {'text': 'house and the suspense is so good and', 'start': 315.68, 'duration': 4.64}, {'text': 'then you think oh', 'start': 318.32, 'duration': 3.92}, {'text': 'here it is here it is and then nothing', 'start': 320.32, 'duration': 4.0}, {'text': 'happens and then they go again and they', 'start': 322.24, 'duration': 3.44}, {'text': "build up the suspense as well and it's", 'start': 324.32, 'duration': 3.52}, {'text': "the score that's building it up building", 'start': 325.68, 'duration': 3.359}, {'text': "up and you're waiting and you're on the", 'start': 327.84, 'duration': 2.4}, {'text': 'edge of your seat', 'start': 329.039, 'duration': 3.121}, {'text': 'that stuff is awesome and i love to see', 'start': 330.24, 'duration': 3.84}, {'text': 'that in these type of films what it', 'start': 332.16, 'duration': 4.479}, {'text': 'really was evident in this film to me', 'start': 334.08, 'duration': 5.2}, {'text': 'was they were more focused on the story', 'start': 336.639, 'duration': 4.721}, {'text': 'element and humanizing and creating this', 'start': 339.28, 'duration': 4.16}, {'text': 'great character', 'start': 341.36, 'duration': 5.119}, {'text': 'rather than just', 'start': 343.44, 'duration': 3.039}, {'text': 'do some cool ass action movie that', 'start': 347.28, 'duration': 3.84}, {'text': "doesn't make any sense so they really", 'start': 349.52, 'duration': 3.519}, {'text': 'worked on that and i really appreciate', 'start': 351.12, 'duration': 4.32}, {'text': 'them doing that as for the villains of', 'start': 353.039, 'duration': 4.0}, {'text': 'the film', 'start': 355.44, 'duration': 3.44}, {'text': "we don't know too much about them", 'start': 357.039, 'duration': 4.081}, {'text': "we it's", 'start': 358.88, 'duration': 4.08}, {'text': "it's a very one once you figure out what", 'start': 361.12, 'duration': 4.4}, {'text': 'they were doing and everything like that', 'start': 362.96, 'duration': 5.2}, {'text': "you're probably gonna be like oh okay i", 'start': 365.52, 'duration': 3.519}, {'text': 'guess', 'start': 368.16, 'duration': 2.4}, {'text': "that would they'll do but i mean it's", 'start': 369.039, 'duration': 3.44}, {'text': "revenge film it's a revenge film i don't", 'start': 370.56, 'duration': 5.12}, {'text': 'need my villains to be these in-depth', 'start': 372.479, 'duration': 6.16}, {'text': 'story villains of like the vaders the', 'start': 375.68, 'duration': 5.28}, {'text': "thanos's the loki's", 'start': 378.639, 'duration': 3.921}, {'text': "you know i don't need them to be like", 'start': 380.96, 'duration': 4.48}, {'text': 'that i just need them to be', 'start': 382.56, 'duration': 4.8}, {'text': 'terrible human beings so i can root for', 'start': 385.44, 'duration': 4.08}, {'text': "my hero that's gonna go around and just", 'start': 387.36, 'duration': 4.72}, {'text': "beat him up and do it and that's all i", 'start': 389.52, 'duration': 3.92}, {'text': "ask for and that's what this film", 'start': 392.08, 'duration': 3.679}, {'text': 'delivers i could i could honestly sit', 'start': 393.44, 'duration': 4.08}, {'text': 'down and watch this movie again and', 'start': 395.759, 'duration': 3.121}, {'text': "again and again it's just so", 'start': 397.52, 'duration': 4.399}, {'text': 'entertaining it really is and the whole', 'start': 398.88, 'duration': 5.599}, {'text': 'first first act is them', 'start': 401.919, 'duration': 4.56}, {'text': 'building the character up our main', 'start': 404.479, 'duration': 4.321}, {'text': 'character they build him up and i', 'start': 406.479, 'duration': 4.56}, {'text': 'appreciate that because a lot of these', 'start': 408.8, 'duration': 4.08}, {'text': 'low budget actions can just go straight', 'start': 411.039, 'duration': 4.081}, {'text': 'into it but this one it took its time to', 'start': 412.88, 'duration': 5.36}, {'text': 'build it up and it made it make sense to', 'start': 415.12, 'duration': 5.359}, {'text': 'why certain things were happening and', 'start': 418.24, 'duration': 4.959}, {'text': "that's what i appreciate so overall", 'start': 420.479, 'duration': 4.0}, {'text': 'i actually really enjoyed this film', 'start': 423.199, 'duration': 2.961}, {'text': 'quite a lot is this going to be the best', 'start': 424.479, 'duration': 3.921}, {'text': 'film of the year no is it going to be my', 'start': 426.16, 'duration': 4.159}, {'text': 'top 10 best films of the year', 'start': 428.4, 'duration': 4.799}, {'text': 'no bu

t i tell you what it was fun it was', 'start': 430.319, 'duration': 4.561}, {'text': 'en
tertaining it was exactly what i', 'start': 433.199, 'duration': 3.921}, {'text': 'thoug
ht it was going to be and more it', 'start': 434.88, 'duration': 4.159}, {'text': 'gave
me that extra bit more', 'start': 437.12, 'duration': 3.68}, {'text': 'which i can reall
y really really', 'start': 439.039, 'duration': 4.321}, {'text': 'appreciate and kudos t
o them like they', 'start': 440.8, 'duration': 5.04}, {'text': 'just they gave me a fun
fun action film', 'start': 443.36, 'duration': 4.0}, {'text': "that's all i can ask for
so let me know", 'start': 445.84, 'duration': 2.479}, {'text': 'in the comments below wh
at did you guys', 'start': 447.36, 'duration': 2.239}, {'text': 'think of this movie did
you love did you', 'start': 448.319, 'duration': 2.88}, {'text': 'hate it smash it down
below smash that', 'start': 449.599, 'duration': 2.88}, {'text': "like button click subs
cribe and i'll see", 'start': 451.199, 'duration': 3.12}, {'text': 'you guys next video
until then stay safe', 'start': 452.479, 'duration': 5.481}, {'text': 'and peace out',
'start': 454.319, 'duration': 3.641}]

In [33]:
```python
# Extract the text and concat
youtube_text = " ".join([item['text'].lower() for item in youtube_api])
print(youtube_text)
```

our scene alive is a new action movie that is coming to the big screen for you all to go and check out this movie stars gerard butler and if you guys know his movie you know pretty much what you're gonna get this movie we follow him as he is looking for his wife who has been kidnapped and he's gone to do it by himself it's red bench phil so i'm going to tell you guys my thoughts on this movie let me know in the comments below what do you guys think of this movie or if you haven't seen it what are your thoughts on the trailer are you excited smash it down below smash that like button click subscribe and let's jump into this review of last scene alive okay so last scene alive is a revenge field that's to its core that's what this film is it's a complete and utter revenge film so me i'm a sucker for those revenge fields so i mean we got john wick john wick's a classic revenge film we've got rambo revenge films we've got all these other revenge films and i'm pretty sure most people enjoy them because they're just so much fun and so much entertainment and we like to see the good guys absolutely murder bad guys it's just it's in our dna it's in our nature that's what we want to see when you put in jared butler you know what you're going to get i mean angel has fallen the whole has fallen trilogy you know what you're going to get from this guy he is the modern day liam neeson that's what he is he does these little action films like this these lower budget ones just like liam neeson used to that's what i put these in that in that sort of category this one's no different it's fun it is fun and i tell you what it has a lot more heart than i thought this movie was going to have way more heart it started to pull up my strengths because jared butler's character i actually felt for him they did a fantastic job humanizing him to the point where we don't just see him as this terrible husband we don't see him as this terrible human being we don't he's not even especially skilled or anything like that he's just a regular real estate agent guy that's what he is he does real estate that's all it is but what are they they did so well is they showed us the whole background as we move forward we see the whole background of their relationship and that's what it's really built on we see the ups and downs and we see this guy that just really cares about his wife and when you take that away from him he is willing to do anything and everything to get it back and that's what i like about it they don't go over the top with him doing these insane skills of like a secret agent a secret soldier anything like that he doesn't do that he is just a generic guy and he does what a generic person would do so it's not polished and i respect that because they could have easily wrote it to where he had served or done something and he's like a special agent and that's why he can do all these cool skills or anything like but it's not that's not what it is and they really made it evident that you could actually follow along and understand his motivation to why he was doing it and you could understand exactly where he's coming from and when they showed us all the stuff that happened with him and his character and how broken he is and how he's trying to build it up and he just wants to be this great husband he's willing to do like some of the stuff he's gone through he's willing to still do like be there for her like it's insane that really gets you on his side so whenever he goes around and does kill anyone anything like that you just push it to the side and go yeah justified justified so that really really worked jamie alexander as the wife she was good for the what the whole 10 minutes she was on screen for so she was great the cop i enjoyed him as well what they really did is they did the cop storyline and his storyline as well you know how they call the cops first then the cops take forever so he goes [ __ ] it i'm gonna go do it myself that happens in this as well but i like the cop as well he does a good job the cinematography is great i really enjoy the cinematography gorgeous locations the camera work was outstanding as well this visual effects there is a massive explosion and you can tell that it is quite evidently low budget which is fine because these movies are low budget so you kind of got to expect that the effects aren't going to be like comic book movie level these don't have budgets of 50 plus million these are the low low budgets and i like these low budget action fields they're cool we need movies like this this is just a nice fun time at the theater and it has some suspense to it as well there's a there's a scene where they do it really really well where he's actually going around the house and the suspense is so good and then you think oh here it is here it is and then nothing happens and then they go again and they build up the suspense as well and it's the score that's building it up building up and you're waiting and you're on the edge of your seat that stuff is awesome and i love to see that in these type of films what it really was evident in this film to me was they were more focused on the story element and humanizing and creating this great character rather than just do some cool ass action movie that doesn't make any sense so they really worked on that and i really appreciate them doing that as for the villains of the film we don't know too much about them we it's it's

a very one once you figure out what they were doing and everything like that you're prob
ably gonna be like oh okay i guess that would they'll do but i mean it's revenge film i
t's a revenge film i don't need my villains to be these in–depth story villains of like
the vaders the thanos's the loki's you know i don't need them to be like that i just nee
d them to be terrible human beings so i can root for my hero that's gonna go around and
just beat him up and do it and that's all i ask for and that's what this film delivers i
could i could honestly sit down and watch this movie again and again and again it's just
so entertaining it really is and the whole first first act is them building the characte
r up our main character they build him up and i appreciate that because a lot of these l
ow budget actions can just go straight into it but this one it took its time to build it
up and it made it make sense to why certain things were happening and that's what i appr
eciate so overall i actually really enjoyed this film quite a lot is this going to be th
e best film of the year no is it going to be my top 10 best films of the year no but i t
ell you what it was fun it was entertaining it was exactly what i thought it was going t
o be and more it gave me that extra bit more which i can really really really appreciate
and kudos to them like they just they gave me a fun fun action film that's all i can ask
for so let me know in the comments below what did you guys think of this movie did you l
ove did you hate it smash it down below smash that like button click subscribe and i'll
see you guys next video until then stay safe and peace out

In [34]:
```python
# Tokenization with NLTK's tokenizer
tokens = word_tokenize(youtube_text)
print(tokens)
```

['our', 'scene', 'alive', 'is', 'a', 'new', 'action', 'movie', 'that', 'is', 'coming', 'to', 'the', 'big', 'screen', 'for', 'you', 'all', 'to', 'go', 'and', 'check', 'out', 't his', 'movie', 'stars', 'gerard', 'butler', 'and', 'if', 'you', 'guys', 'know', 'his', 'movie', 'you', 'know', 'pretty', 'much', 'what', 'you', "'re", 'gon', 'na', 'get', 'thi s', 'movie', 'we', 'follow', 'him', 'as', 'he', 'is', 'looking', 'for', 'his', 'wife', 'who', 'has', 'been', 'kidnapped', 'and', 'he', "'s", 'gone', 'to', 'do', 'it', 'by', 'h imself', 'it', "'s", 'red', 'bench', 'phil', 'so', 'i', "'m", 'going', 'to', 'tell', 'yo u', 'guys', 'my', 'thoughts', 'on', 'this', 'movie', 'let', 'me', 'know', 'in', 'the', 'comments', 'below', 'what', 'do', 'you', 'guys', 'think', 'of', 'this', 'movie', 'or', 'if', 'you', 'have', "n't", 'seen', 'it', 'what', 'are', 'your', 'thoughts', 'on', 'th e', 'trailer', 'are', 'you', 'excited', 'smash', 'it', 'down', 'below', 'smash', 'that', 'like', 'button', 'click', 'subscribe', 'and', 'let', "'s", 'jump', 'into', 'this', 'rev iew', 'of', 'last', 'scene', 'alive', 'okay', 'so', 'last', 'scene', 'alive', 'is', 'a', 'revenge', 'field', 'that', "'s", 'to', 'its', 'core', 'that', "'s", 'what', 'this', 'fi lm', 'is', 'it', "'s", 'a', 'complete', 'and', 'utter', 'revenge', 'film', 'so', 'me', 'i', "'m", 'a', 'sucker', 'for', 'those', 'revenge', 'fields', 'so', 'i', 'mean', 'we', 'got', 'john', 'wick', 'john', 'wick', "'s", 'a', 'classic', 'revenge', 'film', 'we', "'ve", 'got', 'rambo', 'revenge', 'films', 'we', "'ve", 'got', 'all', 'these', 'other', 'revenge', 'films', 'and', 'i', "'m", 'pretty', 'sure', 'most', 'people', 'enjoy', 'the m', 'because', 'they', "'re", 'just', 'so', 'much', 'fun', 'and', 'so', 'much', 'enterta inment', 'and', 'we', 'like', 'to', 'see', 'the', 'good', 'guys', 'absolutely', 'murde r', 'bad', 'guys', 'it', "'s", 'just', 'it', "'s", 'in', 'our', 'dna', 'it', "'s", 'in', 'our', 'nature', 'that', "'s", 'what', 'we', 'want', 'to', 'see', 'when', 'you', 'put', 'in', 'jared', 'butler', 'you', 'know', 'what', 'you', "'re", 'going', 'to', 'get', 'i', 'mean', 'angel', 'has', 'fallen', 'the', 'whole', 'has', 'fallen', 'trilogy', 'you', 'kn ow', 'what', 'you', "'re", 'going', 'to', 'get', 'from', 'this', 'guy', 'he', 'is', 'th e', 'modern', 'day', 'liam', 'neeson', 'that', "'s", 'what', 'he', 'is', 'he', 'does', 'these', 'little', 'action', 'films', 'like', 'this', 'these', 'lower', 'budget', 'one s', 'just', 'like', 'liam', 'neeson', 'used', 'to', 'that', "'s", 'what', 'i', 'put', 't hese', 'in', 'that', 'in', 'that', 'sort', 'of', 'category', 'this', 'one', "'s", 'no', 'different', 'it', "'s", 'fun', 'it', 'is', 'fun', 'and', 'i', 'tell', 'you', 'what', 'i t', 'has', 'a', 'lot', 'more', 'heart', 'than', 'i', 'thought', 'this', 'movie', 'was', 'going', 'to', 'have', 'way', 'more', 'heart', 'it', 'started', 'to', 'pull', 'up', 'm y', 'strengths', 'because', 'jared', 'butler', "'s", 'character', 'i', 'actually', 'fel t', 'for', 'him', 'they', 'did', 'a', 'fantastic', 'job', 'humanizing', 'him', 'to', 'th e', 'point', 'where', 'we', 'do', "n't", 'just', 'see', 'him', 'as', 'this', 'terrible', 'husband', 'we', 'do', "n't", 'see', 'him', 'as', 'this', 'terrible', 'human', 'being', 'we', 'do', "n't", 'he', "'s", 'not', 'even', 'especially', 'skilled', 'or', 'anything', 'like', 'that', 'he', "'s", 'just', 'a', 'regular', 'real', 'estate', 'agent', 'guy', 't hat', "'s", 'what', 'he', 'is', 'he', 'does', 'real', 'estate', 'that', "'s", 'all', 'i t', 'is', 'but', 'what', 'are', 'they', 'they', 'did', 'so', 'well', 'is', 'they', 'show ed', 'us', 'the', 'whole', 'background', 'as', 'we', 'move', 'forward', 'we', 'see', 'th e', 'whole', 'background', 'of', 'their', 'relationship', 'and', 'that', "'s", 'what', 'it', "'s", 'really', 'built', 'on', 'we', 'see', 'the', 'ups', 'and', 'downs', 'and', 'we', 'see', 'this', 'guy', 'that', 'just', 'really', 'cares', 'about', 'his', 'wife', 'and', 'when', 'you', 'take', 'that', 'away', 'from', 'him', 'he', 'is', 'willing', 't o', 'do', 'anything', 'and', 'everything', 'to', 'get', 'it', 'back', 'and', 'that', "'s", 'what', 'i', 'like', 'about', 'it', 'they', 'do', "n't", 'go', 'over', 'the', 'to p', 'with', 'him', 'doing', 'these', 'insane', 'skills', 'of', 'like', 'a', 'secret', 'a gent', 'a', 'secret', 'soldier', 'anything', 'like', 'that', 'he', 'does', "n't", 'do', 'that', 'he', 'is', 'just', 'a', 'generic', 'guy', 'and', 'he', 'does', 'what', 'a', 'ge neric', 'person', 'would', 'do', 'so', 'it', "'s", 'not', 'polished', 'and', 'i', 'respe ct', 'that', 'because', 'they', 'could', 'have', 'easily', 'wrote', 'it', 'to', 'where', 'he', 'had', 'served', 'or', 'done', 'something', 'and', 'he', "'s", 'like', 'a', 'speci al', 'agent', 'and', 'that', "'s", 'why', 'he', 'can', 'do', 'all', 'these', 'cool', 'sk ills', 'or', 'anything', 'like', 'but', 'it', "'s", 'not', 'that', "'s", 'not', 'what', 'it', 'is', 'and', 'they', 'really', 'made', 'it', 'evident', 'that', 'you', 'could', 'a ctually', 'follow', 'along', 'and', 'understand', 'his', 'motivation', 'to', 'why', 'h e', 'was', 'doing', 'it', 'and', 'you', 'could', 'understand', 'exactly', 'where', 'he', "'s", 'coming', 'from', 'and', 'when', 'they', 'showed', 'us', 'all', 'the', 'stuff', 't hat', 'happened', 'with', 'him', 'and', 'his', 'character', 'and', 'how', 'broken', 'h e', 'is', 'and', 'how', 'he', "'s", 'trying', 'to', 'build', 'it', 'up', 'and', 'he', 'j

'ust', 'wants', 'to', 'be', 'this', 'great', 'husband', 'he', "'s", 'willing', 'to', 'do', 'like', 'some', 'of', 'the', 'stuff', 'he', "'s", 'gone', 'through', 'he', "'s", 'willing', 'to', 'still', 'do', 'like', 'be', 'there', 'for', 'her', 'like', 'it', "'s", 'insane', 'that', 'really', 'gets', 'you', 'on', 'his', 'side', 'so', 'whenever', 'he', 'goes', 'around', 'and', 'does', 'kill', 'anyone', 'anything', 'like', 'that', 'you', 'just', 'push', 'it', 'to', 'the', 'side', 'and', 'go', 'yeah', 'justified', 'justified', 'so', 'that', 'really', 'really', 'worked', 'jamie', 'alexander', 'as', 'the', 'wife', 'she', 'was', 'good', 'for', 'the', 'what', 'the', 'whole', '10', 'minutes', 'she', 'was', 'on', 'screen', 'for', 'so', 'she', 'was', 'great', 'the', 'cop', 'i', 'enjoyed', 'him', 'as', 'well', 'what', 'they', 'really', 'did', 'is', 'they', 'did', 'the', 'cop', 'story', 'line', 'and', 'his', 'storyline', 'as', 'well', 'you', 'know', 'how', 'they', 'call', 'the', 'cops', 'first', 'then', 'the', 'cops', 'take', 'forever', 'so', 'he', 'goes', '[', '__', ']', 'it', 'i', "'m", 'gon', 'na', 'go', 'do', 'it', 'myself', 'that', 'happens', 'in', 'this', 'as', 'well', 'but', 'i', 'like', 'the', 'cop', 'as', 'well', 'he', 'does', 'a', 'good', 'job', 'the', 'cinematography', 'is', 'great', 'i', 'really', 'enjoy', 'the', 'cinematography', 'gorgeous', 'locations', 'the', 'camera', 'work', 'was', 'outstanding', 'as', 'well', 'this', 'visual', 'effects', 'there', 'is', 'a', 'massive', 'explosion', 'and', 'you', 'can', 'tell', 'that', 'it', 'is', 'quite', 'evidently', 'low', 'budget', 'which', 'is', 'fine', 'because', 'these', 'movies', 'are', 'low', 'budget', 'so', 'you', 'kind', 'of', 'got', 'to', 'expect', 'that', 'the', 'effects', 'are', "n't", 'going', 'to', 'be', 'like', 'comic', 'book', 'movie', 'level', 'these', 'do', "n't", 'have', 'budgets', 'of', '50', 'plus', 'million', 'these', 'are', 'the', 'low', 'low', 'budgets', 'and', 'i', 'like', 'these', 'low', 'budget', 'action', 'fields', 'they', "'re", 'cool', 'we', 'need', 'movies', 'like', 'this', 'this', 'is', 'just', 'a', 'nice', 'fun', 'time', 'at', 'the', 'theater', 'and', 'it', 'has', 'some', 'suspense', 'to', 'it', 'as', 'well', 'there', "'s", 'a', 'there', "'s", 'a', 'scene', 'where', 'they', 'do', 'it', 'really', 'really', 'well', 'where', 'he', "'s", 'actually', 'going', 'around', 'the', 'house', 'and', 'the', 'suspense', 'is', 'so', 'good', 'and', 'then', 'you', 'think', 'oh', 'here', 'it', 'is', 'here', 'it', 'is', 'and', 'then', 'nothing', 'happens', 'and', 'then', 'they', 'go', 'again', 'and', 'they', 'build', 'up', 'the', 'suspense', 'as', 'well', 'and', 'it', "'s", 'the', 'score', 'that', "'s", 'building', 'it', 'up', 'building', 'up', 'and', 'you', "'re", 'waiting', 'and', 'you', "'re", 'on', 'the', 'edge', 'of', 'your', 'seat', 'that', 'stuff', 'is', 'awesome', 'and', 'i', 'love', 'to', 'see', 'that', 'in', 'these', 'type', 'of', 'films', 'what', 'it', 'really', 'was', 'evident', 'in', 'this', 'film', 'to', 'me', 'was', 'they', 'were', 'more', 'focused', 'on', 'the', 'story', 'element', 'and', 'humanizing', 'and', 'creating', 'this', 'great', 'character', 'rather', 'than', 'just', 'do', 'some', 'cool', 'ass', 'action', 'movie', 'that', 'does', "n't", 'make', 'any', 'sense', 'so', 'they', 'really', 'worked', 'on', 'that', 'and', 'i', 'really', 'appreciate', 'them', 'doing', 'that', 'as', 'for', 'the', 'villains', 'of', 'the', 'film', 'we', 'do', "n't", 'know', 'too', 'much', 'about', 'them', 'we', 'it', "'s", 'it', "'s", 'a', 'very', 'one', 'once', 'you', 'figure', 'out', 'what', 'they', 'were', 'doing', 'and', 'everything', 'like', 'that', 'you', "'re", 'probably', 'gon', 'na', 'be', 'like', 'oh', 'okay', 'i', 'guess', 'that', 'would', 'they', "'ll", 'do', 'but', 'i', 'mean', 'it', "'s", 'revenge', 'film', 'it', "'s", 'a', 'revenge', 'film', 'i', 'do', "n't", 'need', 'my', 'villains', 'to', 'be', 'these', 'in-depth', 'story', 'villains', 'of', 'like', 'the', 'vaders', 'the', 'thanos', "'s", 'the', 'loki', "'s", 'you', 'know', 'i', 'do', "n't", 'need', 'them', 'to', 'be', 'like', 'that', 'i', 'just', 'need', 'them', 'to', 'be', 'terrible', 'human', 'beings', 'so', 'i', 'can', 'root', 'for', 'my', 'hero', 'that', "'s", 'gon', 'na', 'go', 'around', 'and', 'just', 'beat', 'him', 'up', 'and', 'do', 'it', 'and', 'that', "'s", 'all', 'i', 'ask', 'for', 'and', 'that', "'s", 'what', 'this', 'film', 'delivers', 'i', 'could', 'i', 'could', 'honestly', 'sit', 'down', 'and', 'watch', 'this', 'movie', 'again', 'and', 'again', 'and', 'again', 'it', "'s", 'just', 'so', 'entertaining', 'it', 'really', 'is', 'and', 'the', 'whole', 'first', 'first', 'act', 'is', 'them', 'building', 'the', 'character', 'up', 'our', 'main', 'character', 'they', 'build', 'him', 'up', 'and', 'i', 'appreciate', 'that', 'because', 'a', 'lot', 'of', 'these', 'low', 'budget', 'actions', 'can', 'just', 'go', 'straight', 'into', 'it', 'but', 'this', 'one', 'it', 'took', 'its', 'time', 'to', 'build', 'it', 'up', 'and', 'it', 'made', 'it', 'make', 'sense', 'to', 'why', 'certain', 'things', 'were', 'happening', 'and', 'that', "'s", 'what', 'i', 'appreciate', 'so', 'overall', 'i', 'actually', 'really', 'enjoyed', 'this', 'film', 'quite', 'a', 'lot', 'is', 'this', 'going', 'to', 'be', 'the', 'best', 'film', 'of', 'the', 'year', 'no', 'is', 'it', 'going', 'to', 'be', 'my', 'top', '10', 'best', 'films', 'of', 'the', 'year', 'no', 'bu

t', 'i', 'tell', 'you', 'what', 'it', 'was', 'fun', 'it', 'was', 'entertaining', 'it',
'was', 'exactly', 'what', 'i', 'thought', 'it', 'was', 'going', 'to', 'be', 'and', 'mor
e', 'it', 'gave', 'me', 'that', 'extra', 'bit', 'more', 'which', 'i', 'can', 'really',
'really', 'really', 'appreciate', 'and', 'kudos', 'to', 'them', 'like', 'they', 'just',
'they', 'gave', 'me', 'a', 'fun', 'fun', 'action', 'film', 'that', "'s", 'all', 'i', 'ca
n', 'ask', 'for', 'so', 'let', 'me', 'know', 'in', 'the', 'comments', 'below', 'what',
'did', 'you', 'guys', 'think', 'of', 'this', 'movie', 'did', 'you', 'love', 'did', 'yo
u', 'hate', 'it', 'smash', 'it', 'down', 'below', 'smash', 'that', 'like', 'button', 'cl
ick', 'subscribe', 'and', 'i', "'ll", 'see', 'you', 'guys', 'next', 'video', 'until', 't
hen', 'stay', 'safe', 'and', 'peace', 'out']

In [35]:
```python
# Count tokens frequencies
freq = nltk.FreqDist(tokens)
for key, val in freq.items():
  print(str(key) + ':' + str(val))
```

our:4
scene:4
alive:3
is:29
a:23
new:1
action:5
movie:11
that:46
coming:2
to:36
the:45
big:1
screen:2
for:11
you:33
all:7
go:7
and:61
check:1
out:3
this:26
stars:1
gerard:1
butler:3
if:2
guys:7
know:9
his:7
pretty:2
much:4
what:25
're:8
gon:4
na:4
get:4
we:16
follow:2
him:11
as:13
he:28
looking:1
wife:3
who:1
has:5
been:1
kidnapped:1
's:51
gone:2
do:21
it:56
by:1
himself:1
red:1
bench:1
phil:1
so:19
i:35
'm:4
going:9
tell:4
my:5

thoughts:2
on:8
let:3
me:6
in:10
comments:2
below:4
think:3
of:16
or:4
have:4
n't:12
seen:1
are:6
your:2
trailer:1
excited:1
smash:4
down:3
like:24
button:2
click:2
subscribe:2
jump:1
into:2
review:1
last:2
okay:2
revenge:8
field:1
its:2
core:1
film:11
complete:1
utter:1
sucker:1
those:1
fields:2
mean:3
got:4
john:2
wick:2
classic:1
've:2
rambo:1
films:5
these:13
other:1
sure:1
most:1
people:1
enjoy:2
them:7
because:5
they:23
just:16
fun:7
entertainment:1
see:9
good:4
absolutely:1
murder:1

bad:1
dna:1
nature:1
want:1
when:3
put:2
jared:2
angel:1
fallen:2
whole:5
trilogy:1
from:3
guy:4
modern:1
day:1
liam:2
neeson:2
does:7
little:1
lower:1
budget:5
ones:1
used:1
sort:1
category:1
one:3
no:3
different:1
lot:3
more:5
heart:2
than:2
thought:2
was:12
way:1
started:1
pull:1
up:9
strengths:1
character:5
actually:4
felt:1
did:7
fantastic:1
job:2
humanizing:2
point:1
where:5
terrible:3
husband:2
human:2
being:1
not:4
even:1
especially:1
skilled:1
anything:5
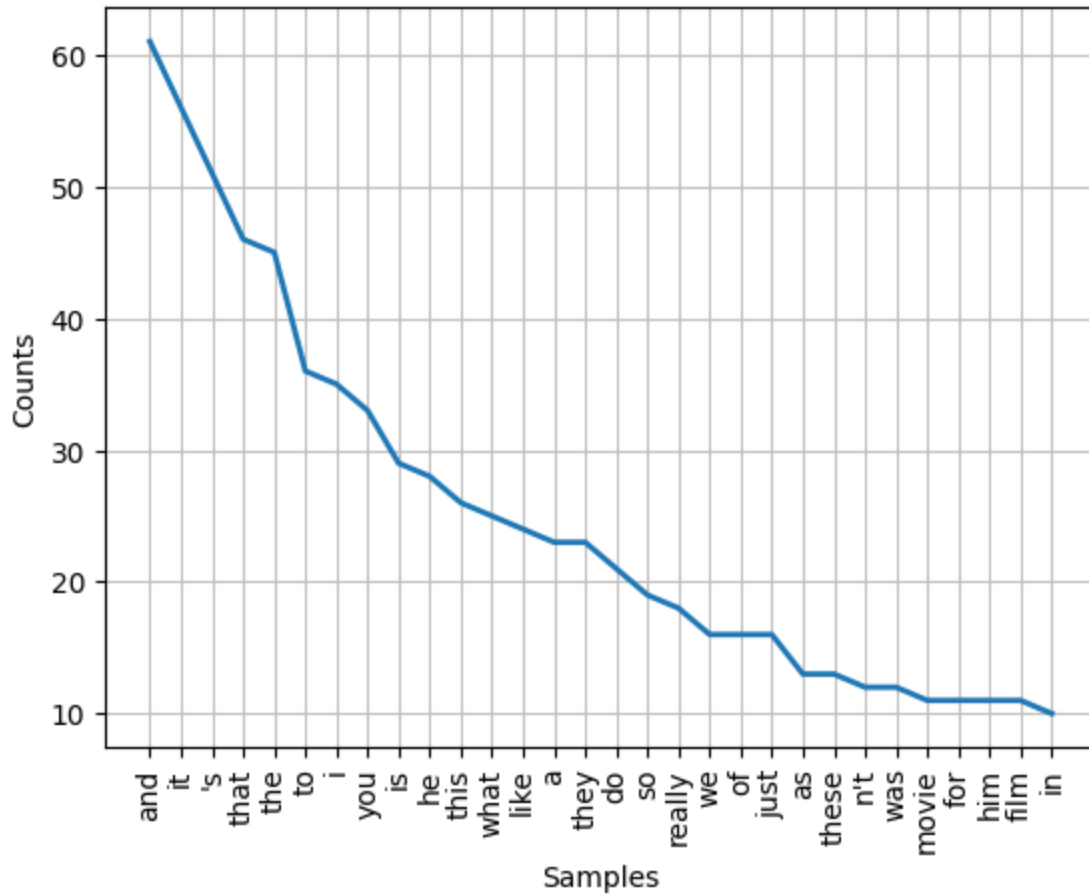regular:1
real:2
estate:2
agent:3
but:6

well:9
showed:2
us:2
background:2
move:1
forward:1
their:1
relationship:1
really:18
built:1
ups:1
downs:1
cares:1
about:3
take:2
away:1
willing:3
everything:2
back:1
over:1
top:2
with:2
doing:4
insane:2
skills:2
secret:2
soldier:1
generic:2
person:1
would:2
polished:1
respect:1
could:5
easily:1
wrote:1
had:1
served:1
done:1
something:1
special:1
why:3
can:6
cool:3
made:2
evident:2
along:1
understand:2
motivation:1
exactly:2
stuff:3
happened:1
how:3
broken:1
trying:1
build:4
wants:1
be:10
great:4
some:3
through:1
still:1
there:4

her:1
gets:1
side:2
whenever:1
goes:2
around:3
kill:1
anyone:1
push:1
yeah:1
justified:2
worked:2
jamie:1
alexander:1
she:3
10:2
minutes:1
cop:3
enjoyed:2
storyline:2
call:1
cops:2
first:3
then:5
forever:1
[:1
__:1
]:1
myself:1
happens:2
cinematography:2
gorgeous:1
locations:1
camera:1
work:1
outstanding:1
visual:1
effects:2
massive:1
explosion:1
quite:2
evidently:1
low:6
which:2
fine:1
movies:2
kind:1
expect:1
comic:1
book:1
level:1
budgets:2
50:1
plus:1
million:1
need:4
nice:1
time:2
at:1
theater:1
suspense:3
house:1

oh:2
here:2
nothing:1
again:4
score:1
building:3
waiting:1
edge:1
seat:1
awesome:1
love:2
type:1
were:3
focused:1
story:2
element:1
creating:1
rather:1
ass:1
make:2
any:1
sense:2
appreciate:4
villains:3
too:1
very:1
once:1
figure:1
probably:1
guess:1
'll:2
in-depth:1
vaders:1
thanos:1
loki:1
beings:1
root:1
hero:1
beat:1
ask:2
delivers:1
honestly:1
sit:1
watch:1
entertaining:2
act:1
main:1
actions:1
straight:1
took:1
certain:1
things:1
happening:1
overall:1
best:2
year:2
gave:2
extra:1
bit:1
kudos:1
hate:1
next:1

```
video:1
until:1
stay:1
safe:1
peace:1
```

In [36]: 
```python
# Draw diagram before normalization
freq.plot(30, cumulative=False)
```



Out[36]: `<AxesSubplot: xlabel='Samples', ylabel='Counts'>`

In [37]: 
```python
# List all Stop Words
nltk.download('stopwords')
from nltk.corpus import stopwords
stopwords.words('english')
```

```
[nltk_data] Downloading package stopwords to
[nltk_data]     /Users/sanjib.paudel/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

```
Out[37]: ['i',
          'me',
          'my',
          'myself',
          'we',
          'our',
          'ours',
          'ourselves',
          'you',
          "you're",
          "you've",
          "you'll",
          "you'd",
          'your',
          'yours',
          'yourself',
          'yourselves',
          'he',
          'him',
          'his',
          'himself',
          'she',
          "she's",
          'her',
          'hers',
          'herself',
          'it',
          "it's",
          'its',
          'itself',
          'they',
          'them',
          'their',
          'theirs',
          'themselves',
          'what',
          'which',
          'who',
          'whom',
          'this',
          'that',
          "that'll",
          'these',
          'those',
          'am',
          'is',
          'are',
          'was',
          'were',
          'be',
          'been',
          'being',
          'have',
          'has',
          'had',
          'having',
          'do',
          'does',
          'did',
          'doing',
          'a',
```

```
    'an',
    'the',
    'and',
    'but',
    'if',
    'or',
    'because',
    'as',
    'until',
    'while',
    'of',
    'at',
    'by',
    'for',
    'with',
    'about',
    'against',
    'between',
    'into',
    'through',
    'during',
    'before',
    'after',
    'above',
    'below',
    'to',
    'from',
    'up',
    'down',
    'in',
    'out',
    'on',
    'off',
    'over',
    'under',
    'again',
    'further',
    'then',
    'once',
    'here',
    'there',
    'when',
    'where',
    'why',
    'how',
    'all',
    'any',
    'both',
    'each',
    'few',
    'more',
    'most',
    'other',
    'some',
    'such',
    'no',
    'nor',
    'not',
    'only',
    'own',
    'same',
    'so',
```

```
                  'than',
                  'too',
                  'very',
                  's',
                  't',
                  'can',
                  'will',
                  'just',
                  'don',
                  "don't",
                  'should',
                  "should've",
                  'now',
                  'd',
                  'll',
                  'm',
                  'o',
                  're',
                  've',
                  'y',
                  'ain',
                  'aren',
                  "aren't",
                  'couldn',
                  "couldn't",
                  'didn',
                  "didn't",
                  'doesn',
                  "doesn't",
                  'hadn',
                  "hadn't",
                  'hasn',
                  "hasn't",
                  'haven',
                  "haven't",
                  'isn',
                  "isn't",
                  'ma',
                  'mightn',
                  "mightn't",
                  'mustn',
                  "mustn't",
                  'needn',
                  "needn't",
                  'shan',
                  "shan't",
                  'shouldn',
                  "shouldn't",
                  'wasn',
                  "wasn't",
                  'weren',
                  "weren't",
                  'won',
                  "won't",
                  'wouldn',
                  "wouldn't"]
```

In [38]:
```python
# Remove stopwords
clean_tokens = tokens[:]
sr = stopwords.words('english')
for token in tokens:
```

```
            if token in sr:
                clean_tokens.remove(token)
```

In [39]:
```
# Use regex to select only words with three or more characters
clean_text = " ".join(clean_tokens)
re_pattern = r'[a-z][a-z][a-z]+'
clean_tokens = re.findall(re_pattern, clean_text)
print(clean_tokens)
```

['scene', 'alive', 'new', 'action', 'movie', 'coming', 'big', 'screen', 'check', 'movie', 'stars', 'gerard', 'butler', 'guys', 'know', 'movie', 'know', 'pretty', 'much', 'gon', 'get', 'movie', 'follow', 'looking', 'wife', 'kidnapped', 'gone', 'red', 'bench', 'phil', 'going', 'tell', 'guys', 'thoughts', 'movie', 'let', 'know', 'comments', 'guys', 'think', 'movie', 'seen', 'thoughts', 'trailer', 'excited', 'smash', 'smash', 'like', 'button', 'click', 'subscribe', 'let', 'jump', 'review', 'last', 'scene', 'alive', 'okay', 'last', 'scene', 'alive', 'revenge', 'field', 'core', 'film', 'complete', 'utter', 'revenge', 'film', 'sucker', 'revenge', 'fields', 'mean', 'got', 'john', 'wick', 'john', 'wick', 'classic', 'revenge', 'film', 'got', 'rambo', 'revenge', 'films', 'got', 'revenge', 'films', 'pretty', 'sure', 'people', 'enjoy', 'much', 'fun', 'much', 'entertainment', 'like', 'see', 'good', 'guys', 'absolutely', 'murder', 'bad', 'guys', 'dna', 'nature', 'want', 'see', 'put', 'jared', 'butler', 'know', 'going', 'get', 'mean', 'angel', 'fallen', 'whole', 'fallen', 'trilogy', 'know', 'going', 'get', 'guy', 'modern', 'day', 'liam', 'neeson', 'little', 'action', 'films', 'like', 'lower', 'budget', 'ones', 'like', 'liam', 'neeson', 'used', 'put', 'sort', 'category', 'one', 'different', 'fun', 'fun', 'tell', 'lot', 'heart', 'thought', 'movie', 'going', 'way', 'heart', 'started', 'pull', 'strengths', 'jared', 'butler', 'character', 'actually', 'felt', 'fantastic', 'job', 'humanizing', 'point', 'see', 'terrible', 'husband', 'see', 'terrible', 'human', 'even', 'especially', 'skilled', 'anything', 'like', 'regular', 'real', 'estate', 'agent', 'guy', 'real', 'estate', 'well', 'showed', 'whole', 'background', 'move', 'forward', 'see', 'whole', 'background', 'relationship', 'really', 'built', 'see', 'ups', 'downs', 'see', 'guy', 'really', 'cares', 'wife', 'take', 'away', 'willing', 'anything', 'everything', 'get', 'back', 'like', 'top', 'insane', 'skills', 'like', 'secret', 'agent', 'secret', 'soldier', 'anything', 'like', 'generic', 'guy', 'generic', 'person', 'would', 'polished', 'respect', 'could', 'easily', 'wrote', 'served', 'done', 'something', 'like', 'special', 'agent', 'cool', 'skills', 'anything', 'like', 'really', 'made', 'evident', 'could', 'actually', 'follow', 'along', 'understand', 'motivation', 'could', 'understand', 'exactly', 'coming', 'showed', 'stuff', 'happened', 'character', 'broken', 'trying', 'build', 'wants', 'great', 'husband', 'willing', 'like', 'stuff', 'gone', 'willing', 'still', 'like', 'like', 'insane', 'really', 'gets', 'side', 'whenever', 'goes', 'around', 'kill', 'anyone', 'anything', 'like', 'push', 'side', 'yeah', 'justified', 'justified', 'really', 'really', 'worked', 'jamie', 'alexander', 'wife', 'good', 'whole', 'minutes', 'screen', 'great', 'cop', 'enjoyed', 'well', 'really', 'cop', 'storyline', 'storyline', 'well', 'know', 'call', 'cops', 'first', 'cops', 'take', 'forever', 'goes', 'gon', 'happens', 'well', 'like', 'cop', 'well', 'good', 'job', 'cinematography', 'great', 'really', 'enjoy', 'cinematography', 'gorgeous', 'locations', 'camera', 'work', 'outstanding', 'well', 'visual', 'effects', 'massive', 'explosion', 'tell', 'quite', 'evidently', 'low', 'budget', 'fine', 'movies', 'low', 'budget', 'kind', 'got', 'expect', 'effects', 'going', 'like', 'comic', 'book', 'movie', 'level', 'budgets', 'plus', 'million', 'low', 'low', 'budgets', 'like', 'low', 'budget', 'action', 'fields', 'cool', 'need', 'movies', 'like', 'nice', 'fun', 'time', 'theater', 'suspense', 'well', 'scene', 'really', 'really', 'well', 'actually', 'going', 'around', 'house', 'suspense', 'good', 'think', 'nothing', 'happens', 'build', 'suspense', 'well', 'score', 'building', 'building', 'waiting', 'edge', 'seat', 'stuff', 'awesome', 'love', 'see', 'type', 'films', 'really', 'evident', 'film', 'focused', 'story', 'element', 'humanizing', 'creating', 'great', 'character', 'rather', 'cool', 'ass', 'action', 'movie', 'make', 'sense', 'really', 'worked', 'really', 'appreciate', 'villains', 'film', 'know', 'much', 'one', 'figure', 'everything', 'like', 'probably', 'gon', 'like', 'okay', 'guess', 'would', 'mean', 'revenge', 'film', 'revenge', 'film', 'need', 'villains', 'depth', 'story', 'villains', 'like', 'vaders', 'thanos', 'loki', 'know', 'need', 'like', 'need', 'terrible', 'human', 'beings', 'root', 'hero', 'gon', 'around', 'beat', 'ask', 'film', 'delivers', 'could', 'could', 'honestly', 'sit', 'watch', 'movie', 'entertaining', 'really', 'whole', 'first', 'first', 'act', 'building', 'character', 'main', 'character', 'build', 'appreciate', 'lot', 'low', 'budget', 'actions', 'straight', 'one', 'took', 'time', 'build', 'made', 'make', 'sense', 'certain', 'things', 'happening', 'appreciate', 'overall', 'actually', 'really', 'enjoyed', 'film', 'quite', 'lot', 'going', 'best', 'film', 'year', 'going', 'top', 'best', 'films', 'year', 'tell', 'fun', 'entertaining', 'exactly', 'thought', 'going', 'gave', 'extra', 'bit', 'really', 'really', 'really', 'appreciate', 'kudos', 'like', 'gave', 'fun', 'fun', 'action', 'film', 'ask', 'let', 'know', 'comments', 'guys', 'think', 'movie', 'love', 'hate', 'smash', 'smash', 'like', 'button', 'click', 'subscribe', 'see', 'guys', 'next', 'video', 'stay', 'safe', 'peace']

```
In [40]:  # Lemmetize words with WordNetLemmatizer
          lemmatizer = WordNetLemmatizer()
          lemma_tokens = [lemmatizer.lemmatize(word) for word in clean_tokens]
          print(lemma_tokens)
```

```
['scene', 'alive', 'new', 'action', 'movie', 'coming', 'big', 'screen', 'check', 'movi
e', 'star', 'gerard', 'butler', 'guy', 'know', 'movie', 'know', 'pretty', 'much', 'gon',
'get', 'movie', 'follow', 'looking', 'wife', 'kidnapped', 'gone', 'red', 'bench', 'phi
l', 'going', 'tell', 'guy', 'thought', 'movie', 'let', 'know', 'comment', 'guy', 'thin
k', 'movie', 'seen', 'thought', 'trailer', 'excited', 'smash', 'smash', 'like', 'butto
n', 'click', 'subscribe', 'let', 'jump', 'review', 'last', 'scene', 'alive', 'okay', 'la
st', 'scene', 'alive', 'revenge', 'field', 'core', 'film', 'complete', 'utter', 'reveng
e', 'film', 'sucker', 'revenge', 'field', 'mean', 'got', 'john', 'wick', 'john', 'wick',
'classic', 'revenge', 'film', 'got', 'rambo', 'revenge', 'film', 'got', 'revenge', 'fil
m', 'pretty', 'sure', 'people', 'enjoy', 'much', 'fun', 'much', 'entertainment', 'like',
'see', 'good', 'guy', 'absolutely', 'murder', 'bad', 'guy', 'dna', 'nature', 'want', 'se
e', 'put', 'jared', 'butler', 'know', 'going', 'get', 'mean', 'angel', 'fallen', 'whol
e', 'fallen', 'trilogy', 'know', 'going', 'get', 'guy', 'modern', 'day', 'liam', 'neeso
n', 'little', 'action', 'film', 'like', 'lower', 'budget', 'one', 'like', 'liam', 'neeso
n', 'used', 'put', 'sort', 'category', 'one', 'different', 'fun', 'fun', 'tell', 'lot',
'heart', 'thought', 'movie', 'going', 'way', 'heart', 'started', 'pull', 'strength', 'ja
red', 'butler', 'character', 'actually', 'felt', 'fantastic', 'job', 'humanizing', 'poin
t', 'see', 'terrible', 'husband', 'see', 'terrible', 'human', 'even', 'especially', 'ski
lled', 'anything', 'like', 'regular', 'real', 'estate', 'agent', 'guy', 'real', 'estat
e', 'well', 'showed', 'whole', 'background', 'move', 'forward', 'see', 'whole', 'backgro
und', 'relationship', 'really', 'built', 'see', 'ups', 'down', 'see', 'guy', 'really',
'care', 'wife', 'take', 'away', 'willing', 'anything', 'everything', 'get', 'back', 'lik
e', 'top', 'insane', 'skill', 'like', 'secret', 'agent', 'secret', 'soldier', 'anythin
g', 'like', 'generic', 'guy', 'generic', 'person', 'would', 'polished', 'respect', 'coul
d', 'easily', 'wrote', 'served', 'done', 'something', 'like', 'special', 'agent', 'coo
l', 'skill', 'anything', 'like', 'really', 'made', 'evident', 'could', 'actually', 'foll
ow', 'along', 'understand', 'motivation', 'could', 'understand', 'exactly', 'coming', 's
howed', 'stuff', 'happened', 'character', 'broken', 'trying', 'build', 'want', 'great',
'husband', 'willing', 'like', 'stuff', 'gone', 'willing', 'still', 'like', 'like', 'insa
ne', 'really', 'get', 'side', 'whenever', 'go', 'around', 'kill', 'anyone', 'anything',
'like', 'push', 'side', 'yeah', 'justified', 'justified', 'really', 'really', 'worked',
'jamie', 'alexander', 'wife', 'good', 'whole', 'minute', 'screen', 'great', 'cop', 'enjo
yed', 'well', 'really', 'cop', 'storyline', 'storyline', 'well', 'know', 'call', 'cop',
'first', 'cop', 'take', 'forever', 'go', 'gon', 'happens', 'well', 'like', 'cop', 'wel
l', 'good', 'job', 'cinematography', 'great', 'really', 'enjoy', 'cinematography', 'gorg
eous', 'location', 'camera', 'work', 'outstanding', 'well', 'visual', 'effect', 'massiv
e', 'explosion', 'tell', 'quite', 'evidently', 'low', 'budget', 'fine', 'movie', 'low',
'budget', 'kind', 'got', 'expect', 'effect', 'going', 'like', 'comic', 'book', 'movie',
'level', 'budget', 'plus', 'million', 'low', 'low', 'budget', 'like', 'low', 'budget',
'action', 'field', 'cool', 'need', 'movie', 'like', 'nice', 'fun', 'time', 'theater', 's
uspense', 'well', 'scene', 'really', 'really', 'well', 'actually', 'going', 'around', 'h
ouse', 'suspense', 'good', 'think', 'nothing', 'happens', 'build', 'suspense', 'well',
'score', 'building', 'building', 'waiting', 'edge', 'seat', 'stuff', 'awesome', 'love',
'see', 'type', 'film', 'really', 'evident', 'film', 'focused', 'story', 'element', 'huma
nizing', 'creating', 'great', 'character', 'rather', 'cool', 'as', 'action', 'movie', 'm
ake', 'sense', 'really', 'worked', 'really', 'appreciate', 'villain', 'film', 'know', 'm
uch', 'one', 'figure', 'everything', 'like', 'probably', 'gon', 'like', 'okay', 'guess',
'would', 'mean', 'revenge', 'film', 'revenge', 'film', 'need', 'villain', 'depth', 'stor
y', 'villain', 'like', 'vaders', 'thanos', 'loki', 'know', 'need', 'like', 'need', 'terr
ible', 'human', 'being', 'root', 'hero', 'gon', 'around', 'beat', 'ask', 'film', 'delive
rs', 'could', 'could', 'honestly', 'sit', 'watch', 'movie', 'entertaining', 'really', 'w
hole', 'first', 'first', 'act', 'building', 'character', 'main', 'character', 'build',
'appreciate', 'lot', 'low', 'budget', 'action', 'straight', 'one', 'took', 'time', 'buil
d', 'made', 'make', 'sense', 'certain', 'thing', 'happening', 'appreciate', 'overall',
'actually', 'really', 'enjoyed', 'film', 'quite', 'lot', 'going', 'best', 'film', 'yea
r', 'going', 'top', 'best', 'film', 'year', 'tell', 'fun', 'entertaining', 'exactly', 't
hought', 'going', 'gave', 'extra', 'bit', 'really', 'really', 'really', 'appreciate', 'k
udos', 'like', 'gave', 'fun', 'fun', 'action', 'film', 'ask', 'let', 'know', 'comment',
'guy', 'think', 'movie', 'love', 'hate', 'smash', 'smash', 'like', 'button', 'click', 's
ubscribe', 'see', 'guy', 'next', 'video', 'stay', 'safe', 'peace']
```

In [41]: `# Count the frequencies of words after normalizations`

```python
lemma_tokens_freq = nltk.FreqDist(lemma_tokens)
for key, val in lemma_tokens_freq.items():
  print(str(key) + ':' + str(val))
```

scene:4
alive:3
new:1
action:6
movie:13
coming:2
big:1
screen:2
check:1
star:1
gerard:1
butler:3
guy:11
know:9
pretty:2
much:4
gon:4
get:5
follow:2
looking:1
wife:3
kidnapped:1
gone:2
red:1
bench:1
phil:1
going:9
tell:4
thought:4
let:3
comment:2
think:3
seen:1
trailer:1
excited:1
smash:4
like:24
button:2
click:2
subscribe:2
jump:1
review:1
last:2
okay:2
revenge:8
field:3
core:1
film:16
complete:1
utter:1
sucker:1
mean:3
got:4
john:2
wick:2
classic:1
rambo:1
sure:1
people:1
enjoy:2
fun:7
entertainment:1

see:9
good:4
absolutely:1
murder:1
bad:1
dna:1
nature:1
want:2
put:2
jared:2
angel:1
fallen:2
whole:5
trilogy:1
modern:1
day:1
liam:2
neeson:2
little:1
lower:1
budget:7
one:4
used:1
sort:1
category:1
different:1
lot:3
heart:2
way:1
started:1
pull:1
strength:1
character:5
actually:4
felt:1
fantastic:1
job:2
humanizing:2
point:1
terrible:3
husband:2
human:2
even:1
especially:1
skilled:1
anything:5
regular:1
real:2
estate:2
agent:3
well:9
showed:2
background:2
move:1
forward:1
relationship:1
really:18
built:1
ups:1
down:1
care:1
take:2

```
away:1
willing:3
everything:2
back:1
top:2
insane:2
skill:2
secret:2
soldier:1
generic:2
person:1
would:2
polished:1
respect:1
could:5
easily:1
wrote:1
served:1
done:1
something:1
special:1
cool:3
made:2
evident:2
along:1
understand:2
motivation:1
exactly:2
stuff:3
happened:1
broken:1
trying:1
build:4
great:4
still:1
side:2
whenever:1
go:2
around:3
kill:1
anyone:1
push:1
yeah:1
justified:2
worked:2
jamie:1
alexander:1
minute:1
cop:5
enjoyed:2
storyline:2
call:1
first:3
forever:1
happens:2
cinematography:2
gorgeous:1
location:1
camera:1
work:1
outstanding:1
visual:1
```

effect:2
massive:1
explosion:1
quite:2
evidently:1
low:6
fine:1
kind:1
expect:1
comic:1
book:1
level:1
plus:1
million:1
need:4
nice:1
time:2
theater:1
suspense:3
house:1
nothing:1
score:1
building:3
waiting:1
edge:1
seat:1
awesome:1
love:2
type:1
focused:1
story:2
element:1
creating:1
rather:1
as:1
make:2
sense:2
appreciate:4
villain:3
figure:1
probably:1
guess:1
depth:1
vaders:1
thanos:1
loki:1
being:1
root:1
hero:1
beat:1
ask:2
delivers:1
honestly:1
sit:1
watch:1
entertaining:2
act:1
main:1
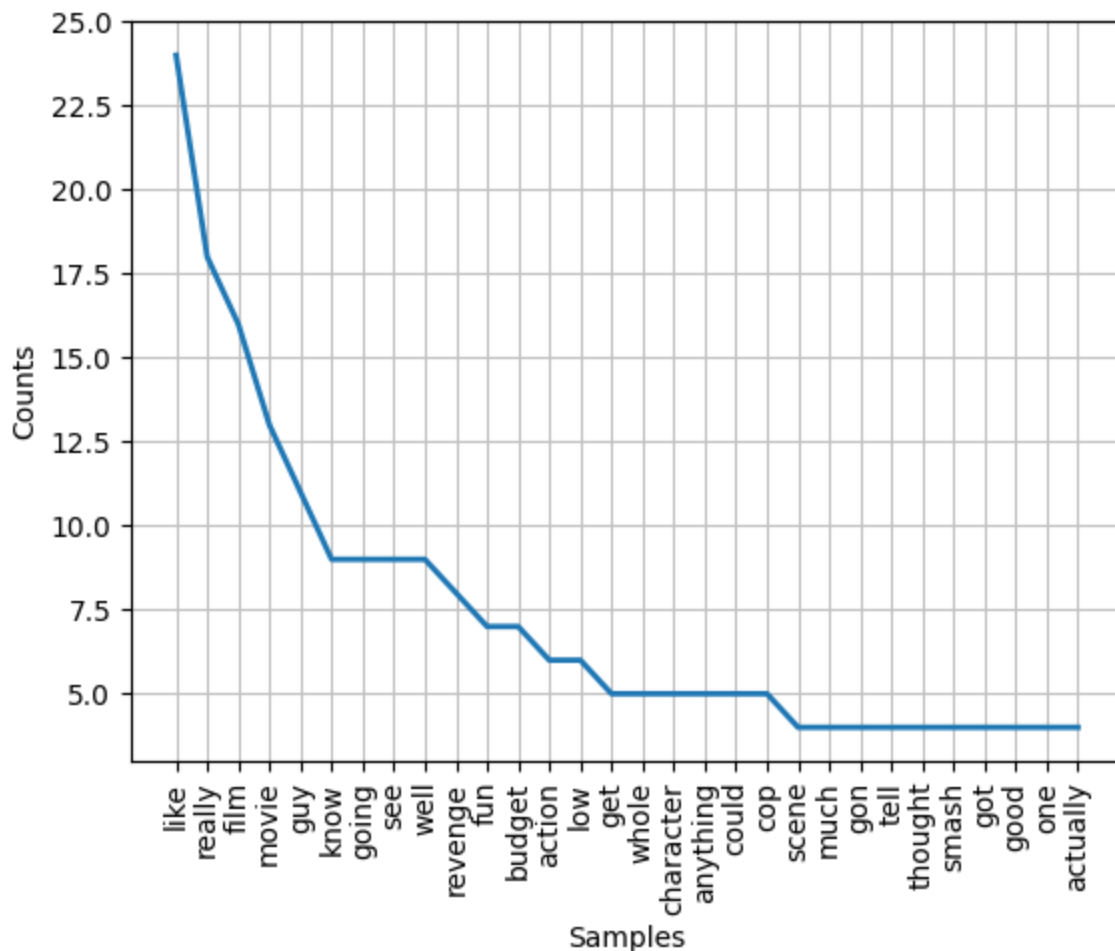straight:1
took:1
certain:1
thing:1

```
happening:1
overall:1
best:2
year:2
gave:2
extra:1
bit:1
kudos:1
hate:1
next:1
video:1
stay:1
safe:1
peace:1
```

In [42]:
```python
# Draw the graph after normalization
lemma_tokens_freq.plot(30, cumulative=False)
```



Out[42]: <AxesSubplot: xlabel='Samples', ylabel='Counts'>

In [43]:
```python
# Methods to clean, create sequence and encode for sentences
# Clean word for sentences
def text_cleaner(text):
    # lower case text
    newString = text.lower()
    newString = re.sub(r"'s\b","",newString)
    # remove punctuations
    newString = re.sub("[^a-zA-Z]", " ", newString)
    long_words=[]
    # remove short word
    for i in newString.split():
        if len(i)>=3:
```

```
                    long_words.append(i)
        return (" ".join(long_words)).strip()

    # Create a sequence os words
    def create_seq(text):
        length = 30
        sequences = list()
        for i in range(length, len(text)):
            # select sequence of tokens
            seq = text[i-length:i+1]
            # store
            sequences.append(seq)
        print('Total Sequences: %d' % len(sequences))
        return sequences

    # Encode sequences
    def encode_seq(seq):
        sequences = list()
        for line in seq:
    #        print([mapping[char] for char in line])
            # integer encode line
            encoded_seq = [mapping[char] for char in line]
            # store
            sequences.append(encoded_seq)
        return sequences
```

In [44]:
```
clean_summary = text_cleaner(youtube_text) # generate clean summary without
seq_summary = create_seq(clean_summary) # generate clean summary sequence
chars = sorted(list(set(clean_summary)))
mapping = dict((c, i) for i, c in enumerate(chars))
encoded_seq = encode_seq(seq_summary) # generate clean summary encode
print(encoded_seq[:10])
```

```
Total Sequences: 5878
[[15, 21, 18, 0, 19, 3, 5, 14, 5, 0, 1, 12, 9, 22, 5, 0, 14, 5, 23, 0, 1, 3, 20, 9, 15,
14, 0, 13, 15, 22, 9], [21, 18, 0, 19, 3, 5, 14, 5, 0, 1, 12, 9, 22, 5, 0, 14, 5, 23, 0,
1, 3, 20, 9, 15, 14, 0, 13, 15, 22, 9, 5], [18, 0, 19, 3, 5, 14, 5, 0, 1, 12, 9, 22, 5,
0, 14, 5, 23, 0, 1, 3, 20, 9, 15, 14, 0, 13, 15, 22, 9, 5, 0], [0, 19, 3, 5, 14, 5, 0,
1, 12, 9, 22, 5, 0, 14, 5, 23, 0, 1, 3, 20, 9, 15, 14, 0, 13, 15, 22, 9, 5, 0, 20], [19,
3, 5, 14, 5, 0, 1, 12, 9, 22, 5, 0, 14, 5, 23, 0, 1, 3, 20, 9, 15, 14, 0, 13, 15, 22, 9,
5, 0, 20, 8], [3, 5, 14, 5, 0, 1, 12, 9, 22, 5, 0, 14, 5, 23, 0, 1, 3, 20, 9, 15, 14, 0,
13, 15, 22, 9, 5, 0, 20, 8, 1], [5, 14, 5, 0, 1, 12, 9, 22, 5, 0, 14, 5, 23, 0, 1, 3, 2
0, 9, 15, 14, 0, 13, 15, 22, 9, 5, 0, 20, 8, 1, 20], [14, 5, 0, 1, 12, 9, 22, 5, 0, 14,
5, 23, 0, 1, 3, 20, 9, 15, 14, 0, 13, 15, 22, 9, 5, 0, 20, 8, 1, 20, 0], [5, 0, 1, 12,
9, 22, 5, 0, 14, 5, 23, 0, 1, 3, 20, 9, 15, 14, 0, 13, 15, 22, 9, 5, 0, 20, 8, 1, 20, 0,
3], [0, 1, 12, 9, 22, 5, 0, 14, 5, 23, 0, 1, 3, 20, 9, 15, 14, 0, 13, 15, 22, 9, 5, 0, 2
0, 8, 1, 20, 0, 3, 15]]
```

In [45]:
```
# Sentence tokenization using Spacy
nlp = spacy.load('en_core_web_sm')
spacy_doc = nlp(clean_summary)
sentences = [sentence.lemma_.lower().strip() for sentence in spacy_doc.sents]
# sentences = [word.lemma_.lower().strip() for word in spacy_doc.sents if word not in st
print("Number of sentences: ", len(sentences))
print(sentences)
```

['our scene alive new action movie that come the big screen for you all and check out this movie star gerard butler and you guy know his movie you know pretty much what you got o get this movie follow he look for his wife who have be kidnap and go himself red bench phil going tell you guy thought this movie let know the comment below what you guy think this movie you haven see what be your thought the trailer be you excited smash down below smash that like button click subscribe and let jump into this review last scene alive okay last scene alive revenge field that its core that what this film complete and utter revenge film sucker for those revenge field mean get john wick john wick classic revenge film get rambo revenge film get all these other revenge film and pretty sure most people enjoy they because they just much fun and much entertainment and like see the good guy absolutely murder bad guy just our dna our nature that what want see when you put jar butler you know what you go get mean angel have fall the whole have fall trilogy you know what you go get from this guy the modern day liam neeson that what do these little action film like this these low budget one just like liam neeson use that what put these that that sort category this one different fun fun and tell you what have lot more heart than think this movie be go have way more heart started pull strength because jared butler character actually feel for he they do fantastic job humanize he the point where don just see he this terrible husband don see he this terrible human being don not even especially skilled anything like that just regular real estate agent guy that what do real estate that all but what be they they do well they show the whole background move forward see the whole background their relationship and that what really build see the up and down and see this guy that just really care about his wife and when you take that away from he willing anything and everything get back and that what like about they don over the top with he do these insane skill like secret agent secret soldier anything like that doesn that just generic guy and do what generic person would not polish and respect that because they could have easily write where have serve do something and like special agent and that why can all these cool skill anything like but not that not what', 'and they really make evident that you could actually follow along and understand his motivation why be do and you could understand exactly where come from and when they show all the stuff that happen with he and his character and how broken and how try build and just want this great husband willing like some the stuff go through willing still like there for she like insane that really get you his side whenever go around and do kill anyone anything like that you just push the side', 'and yeah justify justify that really really work jamie alexander the wife she be good for the what the whole minute she be screen for she be great the cop enjoy he well what they really do they do the cop storyline and his storyline', 'well you know how they call the cop first then the cop take forever go goto myself that happen this', 'well but like the cop well do good job the cinematography', 'great really enjoy the cinematography gorgeous location the camera work be outstanding', 'well this visual effect there massive explosion and you can tell that quite evidently low budget which fine because these movie be low budget you kind get expect that the effect aren go like comic book movie level these don have budget plus million these be the low low budget and like these low budget action field they cool need movie like this this just nice fun time the theater and have some suspense well there there scene where they really really well where actually go around the house and the suspense good', 'and then you think here here and then nothing happen', 'and then they again and they build the suspense well and the score that building building and you wait and you the edge your seat that stuff awesome and love see that these type film what really be evident this film be they be more focus the story element and humanizing and create this great character rather than just some cool ass action movie that doesn make any sense they really work that and really appreciate they do that for the villain the film don know too much about they very one once you figure out what they be do and everything like that you probably goto like okay guess that would they but mean revenge film revenge film don need villain these depth story villain like the vader the thano the loki you know don need they like that just need they terrible human being can root for hero that goto around and just beat he and and that all ask for and that what this film deliver could could honestly sit down and watch this movie again and again and again just entertain really and the whole first first act they build the character our main character they build he and appreciate that because lot these low budget action can just straight into but this one take its time build and make make sense why certain thing be happen and that what appreciate overall actually really enjoy this film quite lot this go the good film the year go top good film the year but tell you what be fun be entertaining be exactly what thought be go and more give t

hat extra bit more which can really really really appreciate and kudo they like they jus
t they give fun fun action film that all can ask for let know the comment below what do
you guy think this movie do you love do you hate smash down below smash that like button
click subscribe and see you guy next video until then stay safe and peace out']

In [55]:
```python
# Create a summary with clean sentences
sentenceValue = {}
for sentence in sentences:
  for word, freq in lemma_tokens_freq.items():
    if word in sentence:
        if sentence in sentenceValue:
            sentenceValue[sentence] += freq
        else :
            sentenceValue[sentence] = freq
sumValues = 0
for sentence in sentenceValue:
    sumValues += sentenceValue[sentence]
average = int(sumValues / len(sentenceValue))

summary = ''
for sentence in sentences:
    if (sentence in sentenceValue) and(sentenceValue[sentence] > (2.9 * average)):
        summary += "" + sentence
print(summary)
```

our scene alive new action movie that come the big screen for you all and check out this
movie star gerard butler and you guy know his movie you know pretty much what you goto g
et this movie follow he look for his wife who have be kidnap and go himself red bench ph
il going tell you guy thought this movie let know the comment below what you guy think t
his movie you haven see what be your thought the trailer be you excited smash down below
smash that like button click subscribe and let jump into this review last scene alive ok
ay last scene alive revenge field that its core that what this film complete and utter r
evenge film sucker for those revenge field mean get john wick john wick classic revenge
film get rambo revenge film get all these other revenge film and pretty sure most people
enjoy they because they just much fun and much entertainment and like see the good guy a
bsolutely murder bad guy just our dna our nature that what want see when you put jar but
ler you know what you go get mean angel have fall the whole have fall trilogy you know w
hat you go get from this guy the modern day liam neeson that what do these little action
film like this these low budget one just like liam neeson use that what put these that t
hat sort category this one different fun fun and tell you what have lot more heart than
think this movie be go have way more heart started pull strength because jared butler ch
aracter actually feel for he they do fantastic job humanize he the point where don just
see he this terrible husband don see he this terrible human being don not even especiall
y skilled anything like that just regular real estate agent guy that what do real estate
that all but what be they they do well they show the whole background move forward see t
he whole background their relationship and that what really build see the up and down an
d see this guy that just really care about his wife and when you take that away from he
willing anything and everything get back and that what like about they don over the top
with he do these insane skill like secret agent secret soldier anything like that doesn
that just generic guy and do what generic person would not polish and respect that becau
se they could have easily write where have serve do something and like special agent and
that why can all these cool skill anything like but not that not what

In [50]:
```python
# textToLower = ""

# # with open("corpus.txt","r") as f1:           # Open our source file
# #     text = f1.read()                          # Read the file
# #     textToLower = text.lower()                # Lower all the capital letter

# words = re.findall("[a-z]+", youtube_text)        # Find all the words and place
# oneString = " ".join(words)                       # Join them into one string
```

```
# pathToFile = "train.txt"                                    # The path we want to store our
# spelling = Spelling(path = pathToFile)                      # Connect the path to the Spelli
# spelling.train(oneString, pathToFile)
```

In [51]:
```
# pathToFile = "train.txt"
# spelling = Spelling(path = pathToFile)
# text = " "


# words = summary.split(" ")
# corrected_summary = " "
# for i in words:
#     corrected_summary = corrected_summary +" "+ spelling.suggest(i)[0][0] # Spell chec

# print(corrected_summary)
```

In [105…
```
# # Create Traning and Validation Set
# # vocabulary size
# vocab = len(mapping)
# dataset = np.array(encoded_seq)
# # create X and y
# X, y = dataset[:,:-1], dataset[:,-1]
# # one hot encode y
# y = to_categorical(y, num_classes=vocab)
# # create train and validation sets
# X_train, X_val, y_train, y_val = train_test_split(X, y, test_size=0.1, random_state=42

# print('Train shape:', X_train.shape, 'Val shape:', X_val.shape)
```

In [106…
```
# # Model Building
# # define model
# model = Sequential()
# model.add(Embedding(vocab, 60, input_length=30, trainable=True))
# model.add(GRU(150, recurrent_dropout=0.1, dropout=0.1))
# model.add(Dense(vocab, activation='softmax'))
# print(model.summary())

# # compile the model
# model.compile(loss='categorical_crossentropy', metrics=['acc'], optimizer='adam')
# # fit the model
# model.fit(X_train, y_train, epochs=100, verbose=1, validation_data=(X_val, y_val))
```

In [62]:
```
# # generate a sequence of characters with a language model
# def generate_seq(model, mapping, seq_length, seed_text, n_chars):
#     in_text = seed_text
#     # generate a fixed number of characters
#     for _ in range(n_chars):
#         # encode the characters as integers
#         encoded = [mapping[char] for char in in_text]
#         # truncate sequences to a fixed length
#         encoded = pad_sequences([encoded], maxlen=seq_length, truncating='pre')
#         # predict character
# #         yhat = model.predict_classes(encoded, verbose=0)
#         yhat = np.argmax(model.predict(encoded), axis=-1)
#         # reverse map integer to character
#         out_char = ''
#         for char, index in mapping.items():
#             if index == yhat:
#                 out_char = char
```

```
#                 break
#        # append to input
#        in_text += char
#     return in_text
```

In [63]:
```
# generated_text = generate_seq(model, mapping, 30, "movie", 15)
# print(generated_text)
```

In [64]:
```
# # Import required libraries
# import torch
# from pytorch_transformers import GPT2Tokenizer, GPT2LMHeadModel

# # Load pre-trained model tokenizer (vocabulary)
# tokenizer = GPT2Tokenizer.from_pretrained('gpt2')

# generated_sentence = ""
# for sentence in sentences:
#     print(sentence.split(" ")[0])
#     text = sentence.split(" ")[0] + " movie"
#     predicted_text = ""
#     while "." not in predicted_text:
#         # Encode a text inputs
#         if predicted_text != "":
#             text = predicted_text

#         indexed_tokens = tokenizer.encode(text)

#         # Convert indexed tokens in a PyTorch tensor
#         tokens_tensor = torch.tensor([indexed_tokens])

#         # Load pre-trained model (weights)
#         model = GPT2LMHeadModel.from_pretrained('gpt2')

#         # Set the model in evaluation mode to deactivate the DropOut modules
#         model.eval()

#         # If you have a GPU, put everything on cuda
#         if(torch.cuda.is_available()):
#             tokens_tensor = tokens_tensor.to('cuda')
#             model.to('cuda')

#         # Predict all tokens
#         with torch.no_grad():
#             outputs = model(tokens_tensor)
#             predictions = outputs[0]

#         # Get the predicted next sub-word
#         predicted_index = torch.argmax(predictions[0, -1, :]).item()
#         predicted_text = tokenizer.decode(indexed_tokens + [predicted_index])
# #            print(predicted_index, predicted_text)
#     generated_sentence += predicted_text

# # Print the predicted word
# print(generated_sentence)
```

In [107…
```
# from nltk.corpus import movie_reviews
# nltk.download('movie_reviews')
```

In [108…
```
# def countFreq(model):
#     for sentence in movie_reviews.sents():
```

```
#            for w1, w2, w3 in trigrams(sentence, pad_right=True, pad_left=True):
#                model[(w1, w2)][w3] += 1
#        return model

# def getProbability(model):
#     for w1_w2 in model:
#         total_count = float(sum(model[w1_w2].values()))
#         for w3 in model[w1_w2]:
#             model[w1_w2][w3] /= total_count

# # Create a placeholder for model
# model = defaultdict(lambda: defaultdict(lambda: 0))
# countFreq(model)
# getProbability(model)
```

In [109…  `# dict(model["movie", "review"])`

In [76]:
```python
import pandas as pd
import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))
```

In [77]:
```python
# Import IMDB datasets
# https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews
df = pd.read_csv("dataset.csv")
df.head()
```

Out[77]:

| | review | sentiment |
|---|---|---|
| **0** | One of the other reviewers has mentioned that ... | positive |
| **1** | A wonderful little production. <br /><br />The... | positive |
| **2** | I thought this was a wonderful way to spend ti... | positive |
| **3** | Basically there's a family where a little boy ... | negative |
| **4** | Petter Mattei's "Love in the Time of Money" is... | positive |

In [78]:  `df['review'].iloc[0]`

Out[78]: "One of the other reviewers has mentioned that after watching just 1 Oz episode you'll be hooked. They are right, as this is exactly what happened with me.<br /><br />The first thing that struck me about Oz was its brutality and unflinching scenes of violence, which set in right from the word GO. Trust me, this is not a show for the faint hearted or timid. This show pulls no punches with regards to drugs, sex or violence. Its is hardcore, in the classic use of the word.<br /><br />It is called OZ as that is the nickname given to the Oswald Maximum Security State Penitentiary. It focuses mainly on Emerald City, an experimental section of the prison where all the cells have glass fronts and face inwards, so privacy is not high on the agenda. Em City is home to many..Aryans, Muslims, gangstas, Latinos, Christians, Italians, Irish and more....so scuffles, death stares, dodgy dealings and shady agreements are never far away.<br /><br />I would say the main appeal of the show is due to the fact that it goes where other shows wouldn't dare. Forget pretty pictures painted for mainstream audiences, forget charm, forget romance...OZ doesn't mess around. The first episode I ever saw struck me as so nasty it was surreal, I couldn't say I was ready for it, but as I watched more, I developed a taste for Oz, and got accustomed to the high levels of graphic violence. Not just violence, but injustice (crooked guards who'll be sold out for a nickel, inmates who'll kill on order and get away with it, well mannered, middle class inmates being turned into prison bitches due to their lack of street skills or prison experience) Watching Oz, you may become comfortable with what is uncomfortable viewing....thats if you can get in touch with your darker side."

In [79]:
```python
# Preprocessing
df['target'] = df['sentiment'].apply(lambda x: 1 if x == 'positive' else 0)
```

In [80]:
```python
df.sample(3)
```

Out[80]:

| | review | sentiment | target |
|---|---|---|---|
| 48842 | I should have trusted my instincts better: No ... | negative | 0 |
| 46597 | Seeing as the vote average was pretty low, and... | positive | 1 |
| 48139 | It seems that some viewers assume that the onl... | negative | 0 |

In [81]:
```python
df.sentiment.value_counts()
```

Out[81]:
```
positive    25000
negative    25000
Name: sentiment, dtype: int64
```

In [82]:
```python
df.shape
```

Out[82]: (50000, 3)

In [84]:
```python
# Lemmatization
# Remove punctuations, next line character, break line character
# Next Line character
def preprocess(text):
    doc = nlp(text)
    preprocessed_text = [token.lemma_ for token in doc if not token.is_punct and
                         token.text != ' ' and token.text != '\n' and token.text != '<br
                         and token.text != '<br/>']
    return " ".join(preprocessed_text)
```

In [85]:
```python
df['preprocessed_review'] = df['review'].apply(preprocess)
```

In [101…
```python
# Split Datasets between training(80%) and testing(20%)
X_train, X_test, y_train, y_test = train_test_split(df['review'], df['target'], test_siz
```
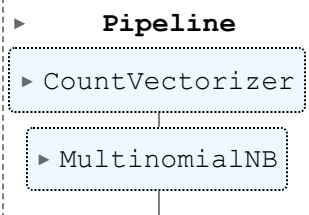
```
In [87]: X_train.shape

Out[87]: (40000,)

In [89]: nbpl = Pipeline([
             ('count_vectorizer', CountVectorizer()),
             ('naive_bayes', MultinomialNB())
         ])

In [90]: # Model Training
         nbpl.fit(X_train, y_train)
```

Out[90]:
```
  ▸       Pipeline
  ┌─────────────────────────┐
  │ ▸ CountVectorizer       │
  └─────────────────────────┘
    ┌─────────────────────┐
    │ ▸ MultinomialNB     │
    └─────────────────────┘
```

```
In [91]: # Testing the model
         y_pred = nbpl.predict(X_test)

In [103…  # nbpl.predict(summary)
          # nbpl.predict(X_test)

In [93]: # Classification report
         from sklearn.metrics import classification_report

         print(classification_report(y_test, y_pred))
                       precision    recall  f1-score   support

                    0       0.83      0.88      0.85      4987
                    1       0.87      0.82      0.84      5013

             accuracy                           0.85     10000
            macro avg       0.85      0.85      0.85     10000
         weighted avg       0.85      0.85      0.85     10000


In [94]: # Random Forest Classifier
         from sklearn.ensemble import RandomForestClassifier

         rfpl = Pipeline([
             ('count_vectorizer', CountVectorizer()),
             ('random_classifier', RandomForestClassifier(n_estimators=50, criterion='entropy'))
         ])

In [95]: # Model training
         rfpl.fit(X_train, y_train)
```
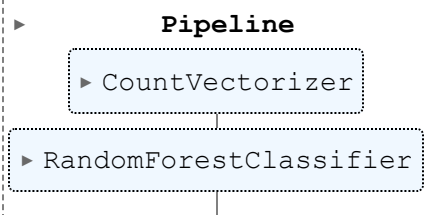
Out[95]:
```
  ▸       Pipeline
  ┌──────────────────────────┐
  │  ▸ CountVectorizer       │
  └──────────────────────────┘
  ┌──────────────────────────┐
  │ ▸ RandomForestClassifier │
  └──────────────────────────┘
```

```
In [96]: # testing the model
```

```
y_pred = rfpl.predict(X_test)
```

In [97]:
```
# Classification report
print(classification_report(y_test, y_pred))
```

```
              precision    recall  f1-score   support

           0       0.85      0.85      0.85      4987
           1       0.85      0.85      0.85      5013

    accuracy                           0.85     10000
   macro avg       0.85      0.85      0.85     10000
weighted avg       0.85      0.85      0.85     10000
```

In [104…
```
# print(y_pred)
# rfpl.predict(summary)
```

In [ ]: