Impact Factor: 19.6
8967:09CX

International Meridian Journal

Leveraging Natural Language Processing and Machine Learning for Enhanced Content Rating

Balaram Yadav Kasula

Dept. of Information Technology

University of The Cumberlands

Williamsburg, KY, USA kramyadav446@gmail.com

Accepted: July 2023

Published: Dec 2023

ABSTRACT: ***In the age of social media, the proliferation of video sharing platforms has attracted a vast user base, with YouTube emerging as one of the most prominent among them, boasting millions of views. The comments section of YouTube videos serves as a treasure trove of information, offering insights that are instrumental in refining content ratings. This paper delves into the realm of sentiment analysis, employing a range of methods and techniques drawn from natural language processing and machine learning. While previous endeavors have explored binary (positive or negative), ternary (including neutral), or even more complex sentiment classifications (such as happy, sad, fearful, surprised, or wrathful), the quest for the most accurate model remains an ongoing challenge. Consequently, this study investigates the application of sentiment analysis to YouTube video comments, with a focus on polarity identification. The research elucidates and categorizes these approaches, offering valuable insights for data mining and sentiment analysis researchers.***

Keywords: *Opinion Mining, Video Sharing Platform, YouTube, YouTube Analysis, Sentiment Analysis, Support Vector Machine,*

INTRODUCTION

The data is collected from the YouTube comments and measures the attitude of the viewer towards the video. Opinion mining or sentiment analysis is useful for analysing the user's opinion based on the large number of text data or comments. Sentimental analysis is the process of searching or tracing the natural language to find patterns or moods of user against a specific topic that is positive, negative, neutral opinions or views. According to the usage statistics, it indicates that 1 billion unique users view 6 billion hours of video each month.

International Meridian Journal

All the data is in the form of text, so textual data needs to be processed. The process starts by first finding out all the words and phrases and thus, a polarity can be assigned. Each comment is assigned one polarity, where polarity means how positive or negative a word or a sentence is and range between -1 to +1. -1 is very negative, +1 is very positive. In this research paper, polarity of the comments is the main concern by Support Vector Machine algorithm.

METHODOLOGY

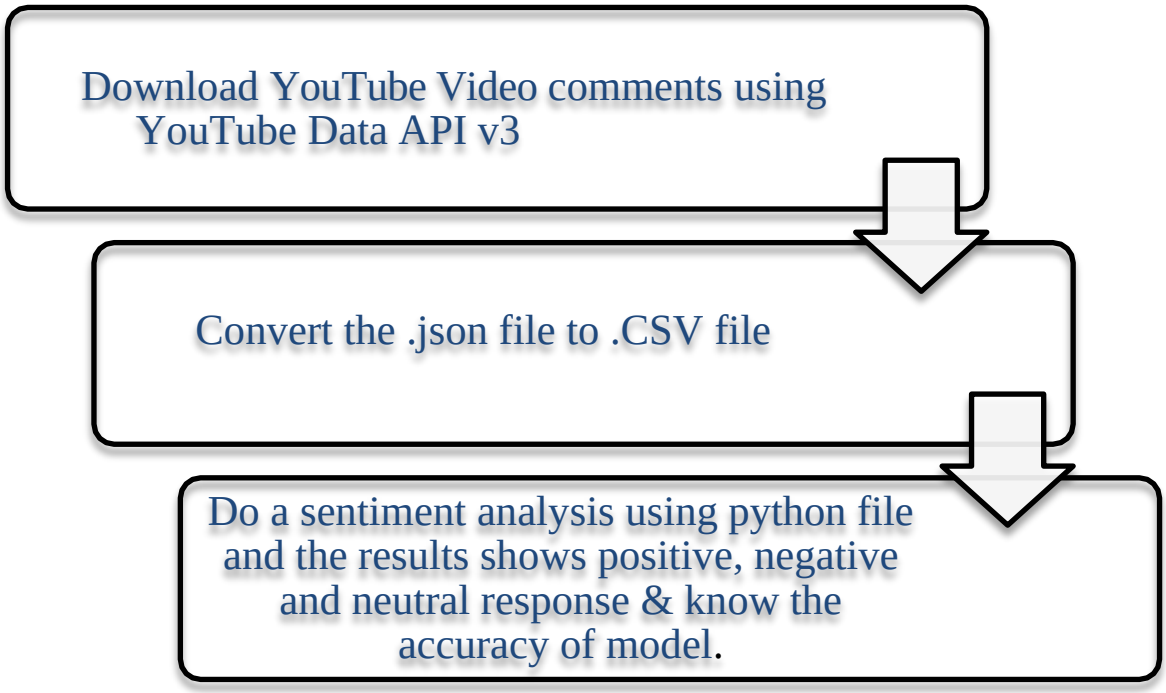

Fig 1. Steps to perform

Support Vector Machine (SVM) is very popular method in research to analyze sentiment. SVM is used in the process of data retrieval and analysis. The process of sentiment analysis is depicted in the below figure.

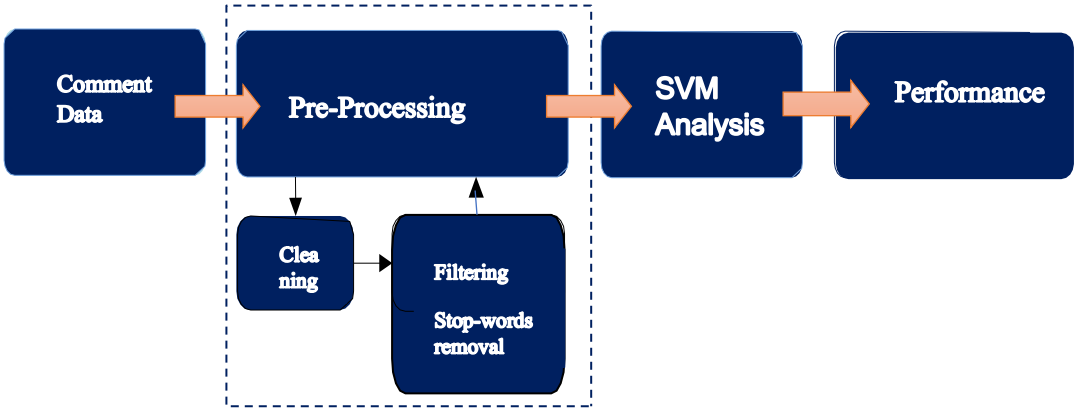

Fig 2. Sentiment Analysis Process

The data was gained by using snipping data and that was taken randomly. Those data were taken and classified into three classes of positive, neutral and negative opinion. The data was extracted and analyzed using SVM method. The data used in this research is the text data of comments and retrieved from YouTube.

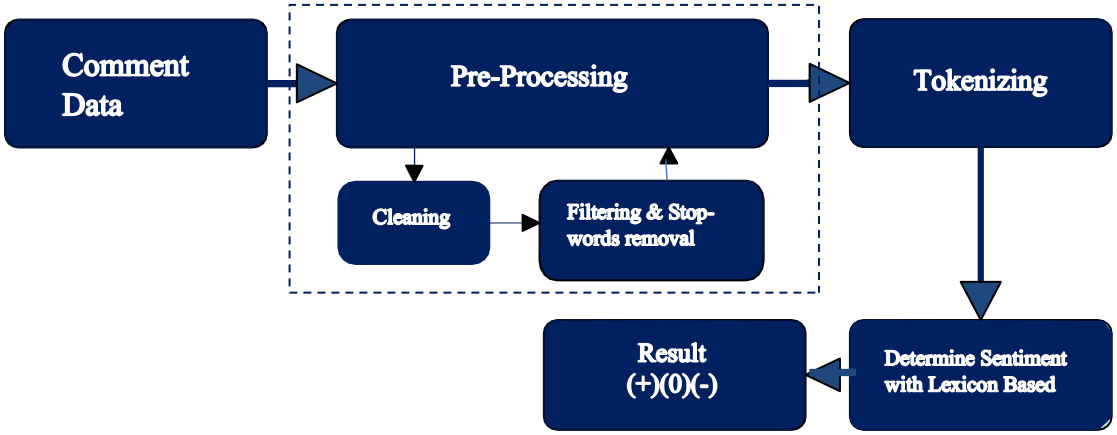

Fig 3. Final Stage of SVM Sentiment Analysis Process

There are several stages involved before analyzing the sentiments, such as

1) Comment Data

Comments from the YouTube were obtained using YouTube Data API v3. The dataset contains many text comments related to the video.

2) Data Pre-Processing

At the data pre-processing stage, there are steps of cleansing and filtering. The cleansing is done to remove links and emojis. The retrieval of important words is done using stop-word removing technique.

3) Tokenizing

The process of removing and deleting some extra words based on the word types intended for punctuation, space, and omitted that are not bound to letters.

4) Determine Sentiment with Lexicon Based

The purpose of lexicon-based method is to determine the sentiments from a sentence. The process of determining is done by summing till *n,* the polarity score of the opinion word, *p* that commented on the feature *f*. The polarity score of an opinion word p will be 1 if the opinion is positive, and -1 if the opinion is negative. Whether the sentence is positive, neutral or negative is determined by the weight of the value in the sentence is done by summing the value of the opinion word that appears. If the value of the opinion word in the sentence is 1, then the value of the sentence is positive, if the value of the opinion word in the sentence is 0, then the value of the sentence is neutral, if the value of the opinion word in the sentence is -1, then the value of the sentence is negative.

Sentiment	*Value*

Positive	1
Neutral	0
Negative	-1

Table 1. Lexicon Method

Based on four basic SVM criteria that is Precision and Recall and performance of the experiments that have been tested to predict the correct and false data. The evaluation of SVM method was done using Confusion Matrix. In confusion matrix, True Positive (TP) is a class which is positive and successfully classified as positive class, True Negative (TN) is a class which is negative and successfully classified as negative, False Positive (FP) is a negative class classified as positive and neutral classes, False Negative (FN) is a positive class and classified as positive and neutral class.

Confusion Matrix

Table 2.

		Actual Value	
		True	False
Predicted Value	True	TP (True Positive) Correct Result	FP (False Positive) Unexpected Result
	False	FN (False Negative) Missing Result	TN (True Negative) Correct Absence of Result

APPENDICES (CODE)

```
# importing libraries
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd


# cleaning the dataset i.e., removing stop words and using stemming technique
import re import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords
```

```
from nltk.stem.porter import PorterStemmer

corpus = []

for i in range(0, 1000):

review = re.sub('[^a-zA-Z]', ' ', dataset['Review'][i])

review = review.lower() review = review.split() ps = PorterStemmer()

all_stopwords = stopwords.words('english') all_stopwords.remove('not')

review = [ps.stem(word) for word in review if not word in set(all_stopwords)]

review = ' '.join(review)

corpus.append(review)


# Creating the bag of words

from sklearn.feature_extraction.text import CountVectorizer

cv = CountVectorizer(max_features = 1500)

X = cv.fit_transform(corpus).toarray()

y = dataset.iloc[:, -1].values


# Splitting the dataset into the Training set and Test set

from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.20, random_state = 0)

# Training the SVM Model

from sklearn.svm import SVC

classifier=SVC(kernel='linear', random_state = 0)

classifier.fit(X_train, y_train)


# Creating the Confusion Matrix of the classification model

from sklearn.metrics import confusion_matrix, accuracy_score

cm = confusion_matrix(y_test, y_pred)

print(cm)
```

accuracy_score(y_test, y_pred)

 # for the visualization of positive sentiments on wordcloud we will store all the comments having polarity 1 in positive_comments

positive_comments=comments[comments['polarity']==1]

RESULTS

The dataset was collected from YouTube Data API v3. The data is taken randomly from the YouTube. The data is divides equally of each class because the problem with the unbalanced data is that the constructed classification has a tendency to ignore the minority class. The data is classified into three classes that is positive, neutral and negative.

The accuracy of SVM method is done by calculating the accuracy, precision and recall performance with confusion matrix method. The evaluation of confusion matrix is done using the following indicators- True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN).

		Actual Value	
		Positive	Negative
Predicted Value	Positive	716	117
	Negative	65	102

Table 3. Confusion Matrix Result

Sentiment Classified Result				
Accuracy (%)	Precision (%)	Recall (%)	TP (%)	TN (%)
83.2%	85.9%	91.6%	91.6%	53.4%

Table 4. The result of YouTube Video Sentiment

CONCLUSION

For understanding or analysing a YouTube video, Support Vector Machine (SVM) is used. The four steps that are involved in this process are Data Comments, Pre-Processing, Tokenizing and Determining the Sentiment using Lexicon. According to Support Vector Machine (SVM) method, the result of the

classification of the True Positive Rate is 91.6% based on the comments. For further development in this research paper, the number of data should be high to achieve more accurate results and conclusions.

REFERENCE

1) Mohammad SM., Zhu X., Kiritchenko S., Martin J (2015) Sentiment, emotion, purpose, and style in electoral tweets.

2) G. Vinodhini., RM. Chandrasekaran., “Sentiment Analysis and Opinion Mining: A Survey”, International Journal of Advanced Research in Computer Science and Software Engineering, Issue Volume 2, 2012.

3) M.Tech., “Methodological Study Of Opinion Mining And Sentiment Analysis Techniques”, International Journal on Soft Computing (IJSC) Vol. 5, No. 1, 2014.

4) Chaitanya Krishna Suryadevara. (2022). UNVEILING COLORS: A K-MEANS APPROACH TO IMAGE-BASED COLOR CLASSIFICATION. International Journal of Innovations in Engineering Research and Technology, 9(9), 47–54. Retrieved from https://repo.ijiert.org/index.php/ijiert/article/view/3577

5) Chaitanya Krishna Suryadevara. (2019). EMOJIFY: CRAFTING PERSONALIZED EMOJIS USING DEEP LEARNING. International Journal of Innovations in Engineering Research and Technology, 6(12), 49–56. Retrieved from https://repo.ijiert.org/index.php/ijiert/article/view/2704

6) Chaitanya Krishna Suryadevara, "Unleashing the Power of Big Data by Transformative Implications and Global Significance of Data-Driven Innovations in the Modern World", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.6, Issue 3, pp.548-554, July 2018, Available at :http://www.ijcrt.org/papers/IJCRT1135233.pdf

7) Chaitanya Krishna Suryadevara, "Transforming Business Operations: Harnessing Artificial Intelligence and Machine Learning in the Enterprise", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.5, Issue 2, pp.931-938, June 2017, Available at :http://www.ijcrt.org/papers/IJCRT1135288.pdf

8) Sobhani P., Mohammad S., Kiritchenko S (2016) Detecting stance in tweets and analyzing its interaction with sentiment, Proceedings of the 5th joint conference on lexical and computational semantics.

9) Moshe Koppel., Jonathan Schler., “The Importance Of Neutral Examples For Learning Sentiment”, Computational Intelligence, Vol. 22, Number 2, 2006.

10) Bo Pang., Lillian Lee., “Opinion Mining and Sentiment Analysis” Vol 2, 2008.

11) Bing Li,. Lei Zhang., “A survey of opinion mining and sentiment analysis” University of Illinois at Chicago, 2012.

12) Suryadevara, Chaitanya Krishna, Feline vs. Canine: A Deep Dive into Image Classification of Cats and Dogs (March 09, 2021). International Research Journal of Mathematics, Engineering and IT, Available at SSRN: https://ssrn.com/abstract=4622112

13) Suryadevara, Chaitanya Krishna, Sparkling Insights: Automated Diamond Price Prediction Using Machine Learning (November 3, 2016). A Journal of Advances in Management IT & Social Sciences, Available at SSRN: https://ssrn.com/abstract=4622110

14) Suryadevara, Chaitanya Krishna, Twitter Sentiment Analysis: Exploring Public Sentiments on Social Media (August 15, 2021). International Journal of Research in Engineering and Applied Sciences, Available at SSRN: https://ssrn.com/abstract=4622111

15) Suryadevara, Chaitanya Krishna, Forensic Foresight: A Comparative Study of Operating System Forensics Tools (July 3, 2022). International Journal of Engineering, Science and Mathematics , Available at SSRN: https://ssrn.com/abstract=4622109

16) Chaitanya krishna Suryadevara. (2023). NOVEL DEVICE TO DETECT FOOD CALORIES USING MACHINE LEARNING. Open Access Repository, 10(9), 52–61. Retrieved from https://oarepo.org/index.php/oa/article/view/3546

17) Chaitanya Krishna Suryadevara, "Exploring the Foundations and Real-World Impact of Artificial Intelligence: Principles, Applications, and Future Directions", International Journal of Creative Research Thoughts (IJCRT), ISSN:2320-2882, Volume.2, Issue 4, pp.22-29, November 2014, Available at :http://www.ijcrt.org/papers/IJCRT1135300.pdf