

Sentimental Analysis on YouTube Scrapped Data

Anusha G¹ and Raghavendra S P²

IV Sem MCA, Department of Master of Computer Applications¹
Assistant Professor, Department of Master of Computer Applications²
JNN College of Engineering, Shimoga, India

Abstract: *In today's world social media plays an important in every one's life in which YouTube plays a vital role. In YouTube anyone can get the required information by specifying the search keyword. One can watch videos and give our opinion based on the content present in that video .YouTube is also a platform which can exhibit our own ideas and show case our talent by which people are getting many employment opportunities. Web scrapping is one of the techniques to scrape the information present in web pages. There are many techniques to scrape YouTube information like using Google API, Selenium python, pafy python, etc. In the proposed work we are scrapping the YouTube video information using paying generating the audience response based on the scrapped information.*

Keywords: Social Media, YouTube, Web scrapping, Opinion Mining, Sentimental Analysis

I. INTRODUCTION

Sentimental analysis is a technique using which we can understand the opinions, sentiments, attitudes, and emotions of the user. It is also known as Opinion Mining (OM). There are mainly four types of sentimental analysis techniques and they are Fine-Grained, Aspect-Based, Emotion Detection, Intent Analysis and there are mainly three types of sentiments they are positive, negative, neutral. It is very useful techniques to get the overall information about the user opinion on certain product, video content, movie reviewed in small amount of time. YouTube has around 122 million active users daily around the world. Every minute more than 500 hours of new content videos are uploaded on YouTube around the world.

In India which rank first in the usage of video has around 467 Million users and they have estimated that by 2025 it may reach 833.03 million users. So there are many ways of opinion are generated based on these contents so the content provider need to be aware about these opinions of the user which will help them to create an improved work. In this research work it is doing runtime scraping on YouTube video comments and generating the list of percentage of sentiment they can see in the comments using the built in packages of python like pay for YouTube video information scrapping, python-comments-scraper to scrape the comments of a particular video and text blob to impose the sentimental analysis on the scraped comments.

II. LITERATURE SURVEY

Several researches and survey has been done on YouTube data scrapping and sentimental analysis .Some of the works has been listed below

Mohd Majid Akhtar et al.,[4] has published the paper Sentiment Analysis on YouTube Comments: A brief study. It uses text blob, a built library to apply sentimental analysis of csv file containing comments.

Viny Christani M et al.,[10] has published paper named COMMENTS SCRAPING APPLICATION FOR REVIEW YOUTUBE CONTENT This paper implemented PHP and HTML programming languages for data mining. Ritika Singh et al.,[5]has published the paper named YOUTUBE COMMENTS SENTIMENT ANALYSIS which contains Naive Bayes, Support Vector Mission, Decision tree, Random Forest, KNN to apply sentimental analysis on YouTube comments. Sowmiya K et al.,[8] has published the paper named Scraping and Analysing YouTube Trending Videos for BI. This paper has implemented Google developer console, YouTube data API and natural language processing to determine the frequency of channels with trending videos.

G Vinodhini et al.,[1] has published paper named Sentiment Analysis and Opinion Mining: A Survey. It is found that the Support Vector Machine and KNN used methodology has outperformed when compared to others. Saukar Anand V et al.,[6] has published paper named An Overview On Web Scraping Techniques And Tools in which it used different

techniques of web scraping and some of the recent tools used for a web scraping. Soujanya Poriya et al.,[7] has published paper named Sentiment Data Flow Analysis by Means of Dynamic Linguistic Patterns. This paper used electronic logic gates algorithm to determine the polarity of each word.

III. PROPOSED METHODOLOGY

In this proposed methodology it is mainly focusing on the extracting the YouTube video information and the comments given by the user for a particular video which is specified by the user through passing the URL of the video.

3.1 Algorithm

- Step1: Enter the keyword to be searched.
- Step2: Pass the keyword as a parameter to the urlopen function.
- Step3: List the video IDs of top 20 videos.
- Step4: Generate URL using these ids.
- Step5: Pass these URL as a parameter to the pafy.new () method.
- Step6: Scrape required attributes like view count, like count, dislike count, title, video length.
- Step7: Store this information in a csv file.
- Step8: Enter the URL to download the comments.
- Step9: Scrape the top 40 comments of that video.
- Step10: Store the comments in a csv file.
- Step11: Apply sentimental analysis.
- Step12: List the count of positive, negative and neutral sentiments.

3.2 Block Diagram

The block diagram gives a better understanding about the proposed work. By just taking just a glance of it we can clearly understand in which step what process is executing. The Figure 1 display the block diagram of the proposed work

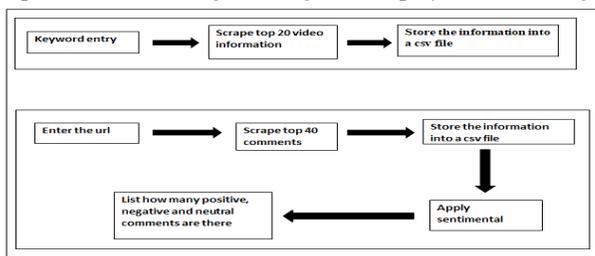


Figure 1: Block Diagram of proposed methodology

1. Ask the users to enter the keyword that they want to search in YouTube. Later it will scrape the top 20 videos information like like count, dislike count, title etc. Then it will store this information in a csv file.
2. Ask the users to enter the URL of the video of which they want to extract the comments. Later those 40 scrapped comments will be stored in a CSV file. The proposed model will apply sentimental analysis on these comments and list the count of positive, negative and neutral comments.

IV. RESULT ANALYSIS

Result analysis is where we analyze the result of each stage of the proposed work like accepting the keyword, storing the videos information, giving the general audience response, accepting the URL, storing the comments, applying sentimental analysis on them.

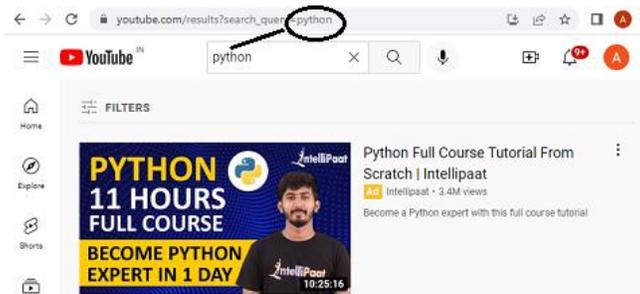


Figure 2: Keyword Searching

In the first stage of the proposed work it will ask the user to enter the keyword based on which the YouTube will list the top 20 videos which can be further analyzed in the next step. As it is visible in the Figure 2 the user need to pass the keyword in the search bar that will be added in the end of the searching URL.

	title	url	viewcount	likecount	dislikecount	Audience_Response
0	Python Tu	https://w	13976619	436163	0	positive
1	Python Tu	https://w	13976619	436163	0	positive
2	Python Tu	https://w	13976619	436163	0	positive
3	Python Tu	https://w	13976619	436163	0	positive
4	Python Tu	https://w	13976619	436163	0	positive
5	Python Tu	https://w	13976619	436163	0	positive
6	Python Tu	https://w	13976619	436163	0	positive
7	Python Tu	https://w	13976619	436163	0	positive
8	Python Tu	https://w	13976619	436164	0	positive
9	Python Tu	https://w	13976619	436164	0	positive
10	Python Tu	https://w	13976619	436164	0	positive
11	Python Tu	https://w	13976619	436164	0	positive
12	Python Tu	https://w	13976619	436164	0	positive
13	Python Tu	https://w	13976619	436164	0	positive
14	Python Tu	https://w	13976619	436164	0	positive
15	Python Tu	https://w	13976619	436164	0	positive
16	Python Tu	https://w	13976619	436164	0	positive
17	Python Tu	https://w	13976619	436164	0	positive
18	Python Tu	https://w	13976619	436164	0	positive
19	Python Tu	https://w	13976619	436166	0	positive

Figure 3: Video Information

The Figure 3 represents the screenshot of the CSV file created after the scrapping of YouTube video information like title, URL, view count, like count, dislike count of each video and the general audience response is generated based on the like and dislike count of the respected video.

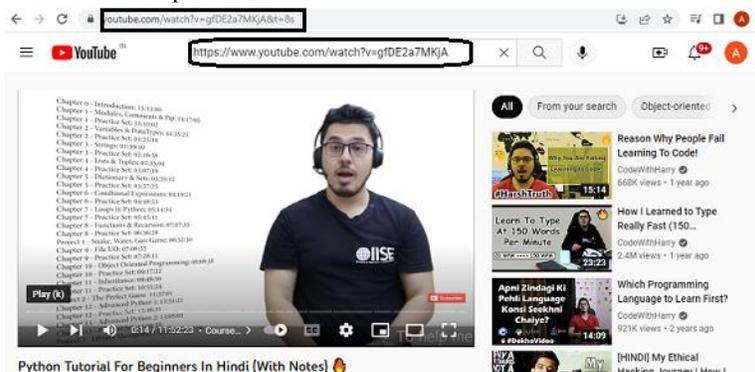


Figure 4: URL Searching

After analysing the general response it will ask the user to enter the URL of a particular video to scrape the top 40 comments which is used for the further analysis. The Figure 4 represent the screenshot of URL passing. After the URL is passed it will automatically open a pop up window that the video will start playing.

	Comment	Likes	Time	UserLink	user
0	[LIMITED STOCK] Buy a cod	175	1 month a	https://www.youtube.com/channel/UC	
1	Dil se shukriya Harry sir ..lc	8.6K	1 year ago	https://w	WRESTLING STAR-WWE
2	7:06:46i have created my fi	15	1 day ago	https://w	Jajati Satpathy
3	6:52:00		3 2 days agc	https://w	Sabnam Laskar
4	1:37:39		1 19 hours a	https://w	Arihant Jain
5	for printing the "object is c		4 2 days agc	https://w	Mitali Sawarkar
6	n = int(input("enter a num		4 2 days agc	https://w	Basic of computer science
7	The guy is working hard ev	1.2K	1 year ago	https://w	Rocker Techs
8	Whenever I thought about	29	8 days agc	https://w	Sahil Jadhav
9	I was trying to learn pytho		28 2 weeks a	https://w	Shashwat Kumar
10	5:13:34		3 1 day ago	https://w	HARSH DESHMUKH
11	Sir I'm in class 8th right no		4 1 day ago	https://w	coderr003
12	5:05:20a = input("Enter use		2 2 days agc	https://w	Noman Khan
13	The only who is providing	226	1 year ago	https://w	Krishnan Kundan
14	I am from Bangladesh and		2 1 day ago	https://w	faiyaz mahmud
15	1:37:06 (day 3)		6 2 weeks a	https://w	Vanshit
16	5:51:41		3 3 days agc	https://w	SHASHANK SINGH
17	9:14:56a function is used fr		3 3 days agc	https://w	Abhijeet Redekar
18	Guys' let's not forget he is	893	1 year ago	https://w	Sayed Abid

Figure 5:Comments received

While watching the video we can scrape the comments and store that in to a CSV file which can be used for further analysis. In that CSV file it contains comment, likes to that particular comment, Time, user link and username of that particular comment. The Figure 5 represents the screenshot of CSV file.

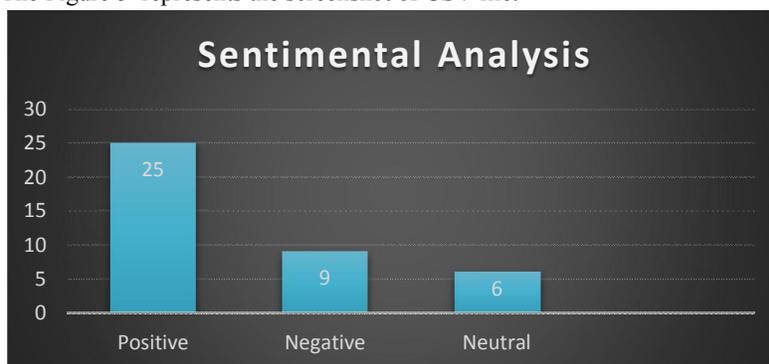


Figure 6: Sentiment Analysis bar chart

It is a bar chart representing how many positive, negative and neutral comments are present in the scrapped comments. As it is shown in Figure 6 there are 25 positive comments, 09 negative comments and 06 neutral comments in the 40 scrapped comments.

V. CONCLUSION

It has been great challenges for the researchers to scrape the comments preprocess them and apply sentimental analysis on them. In this proposed work there is a usage of built-in python libraries which has reduced the middle steps and made the developer to code easily and which has an accuracy of 70%. It can be further improved by identifying the sarcastic comments which can be easily misguidable so that it can provide more accuracy.

ACKNOWLEDGMENT

I would like to express my gratitude internal guide Dr. Raghavendra S P who guided us throughout the project with his insight and knowledge. I would also like to thank our institution, for providing us an opportunity to conduct the project.

REFERENCES

- [1]. G. Vinodhini, RM. Chandrasekaran[2012], Sentiment Analysis and Opinion Mining: A Survey, International journal of Advanced Research in Computer Science and Software Engineering, ISSN:2277 128X, volume 2.
- [2]. <https://www.statista.com/forecasts/1146150/youtube-users-in-india>
- [3]. <https://www.globalmediainsight.com/blog/youtube-users-statistics>
- [4]. Mohd Majid Akhtar[2019], Sentimental Analysis on Youtube Comments: A brief study

- [5]. Ritika Singh, Ayushka Tiwari [2021], Youtube Comments Sentimental Analysis, International Journal of Scientific Research in Engineering and Management(IJSREM),ISSN:2582-3930.
- [6]. Saurkar, Anand V; Pathare, Kedar G; Gode Shweta A. (2018). An Overview On Web Scraping Techniques and Tools. International Journal on Future Revolution in Computer Science & Communication Engineering Volume 4.
- [7]. Soujanya Poria, Erik Cambria, Alenxander Gelbukh, Federic Bisio, Amir Hussain[2015] Sentiment Dataflow Analysis by means of Dynamic Linguistic Patterns.
- [8]. Sowmiya K,Supriya S, R Subhashini[2021], Scarping and Analysing Youtube Trending Vedios for BI,IOS Press.
- [9]. S P Raghavendra, M.J.Adarsh, Shoieb Ahamed and J. Shree Hari, "Estimation of Human Age and Gender Based on LBP Features Using Two Level Decision by SVM" _ K. C. Santosh and B. Gawali (Eds.): RTIP2R 2020, CCIS 1380, pp. 82– 94, 2021. https://doi.org/10.1007/978-981-16-0507-9_8, _ Springer Nature Singapore Pte Ltd. 2021.
- [10]. S P Raghavendra, Ajit Danti, "Recognition of Signature Using Neural Network and Euclidean Distance for Bank Cheque, K. C. Santosh and R. S. Hegadi (Eds.): RTIP2R 2018, CCIS 1037, pp. 228–243, 2019. https://doi.org/10.1007/978-981-13-9187-3_21, Springer Nature Singapore Pte Ltd. 2019
- [11]. TextBlob read blogs. Version v0.15.2.
- [12]. Viny Christanti M, Walda, Tri Sutrisno[2019], Comments Scarping Applications for Review Youtube Content, IOP Conf.Series: Materials Science and Engineering 852.

BIOGRAPHY

- [1]. PYTHON CRASH COURSE E02: A Hands-On, Project-Based Introduction to Programming by Eric Matthes
- [2]. Web Scraping with Python: Collecting Data from the Modern Web Book by Ryan Mitchell
- [3]. Data Analytics using Python by Bharti Motwani.

