Pakistan Journal of Engineering and Technology, PakJET
Multidisciplinary | Peer Reviewed | Open Access
Volume: SI, Number: 01, Pages: 209- 212, Year: 2020

YouTube Video Recommendation Based Multi-lingual Feedback

M. Kashan Akram, M. Abubakar siddique, Mutiullah Jamil, M. Ali Javaid, and Abdullah Haider

Khwaja Freed University of Engineering and Information Technology, Rahim Yar Khan, Pakistan.
Corresponding author: First A. Muhammad Kashan Akram (e-mail: Kashan650@gmail.com).

Abstract: Today YouTube is the first one application when the people need to learn something this is due to the robust service of the YouTube. Every second 500Hours video uploaded on the YouTube. On the one side this huge data is so big edge for YouTube but on the other side this is the big challenge for YouTube to provide efficient and effective search results for each user. People when search on the YouTube they are facing the difficulty to find the best video matched with their required content the only way to find the quality of the video is user feedback in term of comments but we also faced some problem too in comments most important are these two we focus 1.Huge comments 2.Multi-language comments for this problem our proposed method help the user to processes the all comments into the single English language and find the sentiment of the each video category and one the basis of the polarity score we find the best video tutorial and also compare the polarity results with and without our proposed method the results shows the method is effective and efficient.

Keyword: *YouTube, Multi-language comments, Polarity, Sentiment score*

I. INTRODUCTION

YouTube is the first big one video sharing platform today YouTube have 2 billion active users because of its easy and quick access YouTube age is 16 years but its progress and effect in our life shows that we have help if even no one can help us we can get the help from YouTube [1]. Today the YouTube is the platform where the 30,000 hours video new content is uploaded each day and actively 1 billion hours averagely watched in a day [1-5].

Here are the statistics are about the YouTube. The age factor on the YouTube is between 15 years to 25 years but the age 71 is also use the YouTube [1-2]. For the good performance and good interaction YouTube also developed the local versions of the YouTube in the more than 100 countries. 15.2% user are form USA and 8.7% from India and 4.5% from Russia there is 51% of the whole watcher is use the YouTube for do thing he / she never done before 28% user use the YouTube for entertainment or passing the time 19% persons search the product for buying and 19% use the YouTube for knowing the world about events or news more than 5 billion mobile devices the YouTube is installed [3].

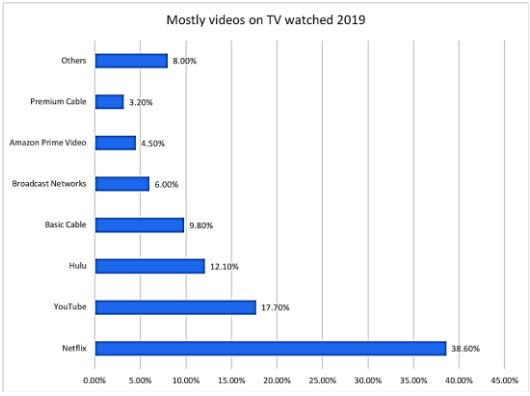

FIGURE 1: Videos watched in 2019

When the user search on the YouTube search engine the YouTube suggest based on the geolocation shown in Fig. 1 and interest of the user and user history this suggestion not the most of time helpful for user after the searching of 10mints the search engine will understand the user requirement and also the user will come to know to search to get the targeted results.

YouTube have many products and enable the user to edit upload and share by YouTube Studio application. YouTube also provides the free video channel making feature from this the many people upload the different unique content from these content people can earn based on the 1000 subscriber and 4000minits watch time is the initial requirements for enabling ad sense option [6-8]. January 2020 YouTube announced that the 37% of the YouTube content is misleading with is the biggest part of the YouTube content. From this misleading content how to can find the leading content for this the best way is comments or the feedback of the videos and we can find the best qualitative content but we also have some problems too in the this scenario some people are not properly feedback on the video content and some people are put the feedback in their own language this also create a new challenge for calculating the positive, negative, and neutral feedback [7].

OUR CONTRIBUTIONS:

1. Find the all language comments.
2. Translate the all languages into one single Language (English).
3. Find the Polarity before translation and After translation.
4. Find the best video based on the proposed method from 5 videos of one category.

Our work is focus on the find the best tutorial video on the 5 videos of the one category and we scrap 4 categories with names below.

1. Machine Learning
2. Data Structure
3. WordPress
4. Python Tutorial

Here are the list of the categories and each category having 5 videos with the same keyword.

II. LITERATURE REVIEW

YouTube ranking done with the help of the polarity score and the people find the video with best average percentage score based on the given range they introduced 4 categories and percentage of the average score which is based on the replies and mostly positive replies are used to rank the video the dataset is collection by opensource ytscrapper.com which is now closed by YouTube [1].
The YouTube comments can be classify in their core concern and study prove that the programmer mostly use the YouTube for to get skills in write, debug and run the code and find the solution for their error faced and in this study 6000 comments are scrapped and collected from 12 tutorial videos about the coding tutorials and they automate the classify the comments based on the machine learning algorithms which is support vector machine with 77% classification accuracy [2]. Qualitative analysis of the YouTube is almost everyone required the machine learning helps the user to find the mood of the user and also can find the spam comments in the feedback on the YouTube [3]. Mood analysis and spam filtering shows the results with reduced false positive 12% and 11% on averagely [4]. Misleading content is the 37% part of the YouTube videos this kind of videos are called clickbait videos such videos are filtered by using machine learning algorithms with collection of 80K videos metadata and the most accurately classify the clickbait videos by using the Random forest with random sampling method [5-7].

YouTube recommendation system also the system of finding the qualitative video content here are two type of recommendation of the video one is when user search in the search bar and select the one listed suggested by the google search box and then this video will be display and more over the YouTube having a feature in the right side suggested videos and this kind of suggestions is called the recommendation both are not the same [8-10].

This study get the relevant information of the user and rank the video plus the other users information about the same video and then rank the video content this methodology is so effective and productive in the study the Deep Neural networks help the user to filter the data [11-12]. Provide the effective recommendation of the video content relevant to your search and user history [13].YouTube recommendation system mostly based on the sentiment analysis or the sentiment score the sentiment analysis means people feedback in term of positive feedback , negative feedback , neutral feedback of the comments and one comment having one sentiment label and the study in Bangla text the one comment can have multi-label sentiment label and they also work on the more strong labels like strongly positive feedback , strongly negative feedback and strongly neutral feedback. They also work on the emotions with five labels on the both models with 3 labels with 66% and 54.24% on the five labels [14].

III. METHODOLOGY

Since the YouTube stop his free API for Comments scrapping now its commercial so we choice the other tools like instant Scrapper, Beautiful Soup to scrapping the YouTube video comments and after collecting the data we process the cleaning data with removing the unknown language comments we got the cleaned .csv of each video comment.
Here are two libraries (Packages) in python and free API for languages detection and translation. Here is our processes model diagram shown in Fig. 2.
Links for libraries:
https://textblob.readthedocs.io/en/dev/
https://pypi.org/project/googletrans/

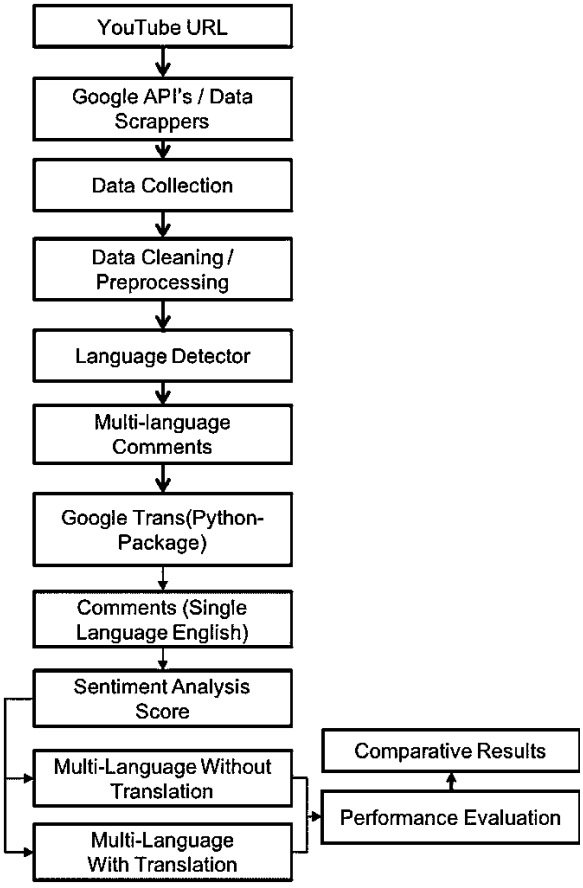

Fig 2: The proposed model

After cleaning data, we put the comments to translate the comments into English even in any language. The GoogleTans [15], and TextBlob [16] helps us to do this there are a lot of free API available but we chose this and translate the comments into one normalized single language which is English. We chose the English because the SentiWord [17] is already build for English and we can find out the polarity of the comments very fast and free. We the sentiment analysis we use the python VaderSentiment [17] package for social media sentiment analysis we use this for polarity calculation and then put each category video to VaderSentiment and then make the label encoding for sentiment analysis and then calculate the final score for the video positive, negative, neutral. since we calculate the both sentiment analysis of the each with and without translation. Here is the translation processes model shown in Fig. 3.

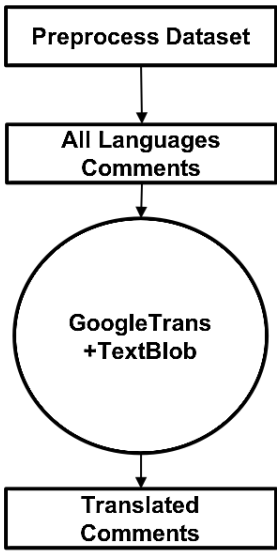

FIGURE 3: Translation process model.

IV. RESULTS AND DISCUSSIONS

We first find the video polarity score without our method and then find the sentiment with our method. Here we are selected the 10 videos with two categories this category are searched by keyword and based on the keyword google shows the results and suggested videos from the google suggested videos we select the first 5 videos for each category and then scrap the comments and apply the method and find the results with and without below graph shown in Fig. 4 is the polarity results of the data structure tutorial.

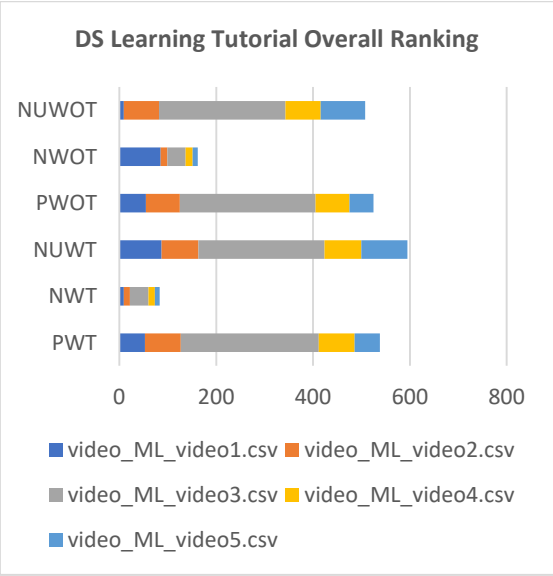

FIGURE 4: DS learning tutorial overall ranking.

The naming convention is shown in the graph.
PWT = Positive Comments with Translation.
NWT = Negative comments with Translation.
NUWT = Neutral comments with Translation.
PWOT = Positive comments without translation.
NWOT = Negative comments without translation.
NUWOT = Neutral comments without comments.

In Fig. 4, the video_3 is best video and ranked in the category even this video low ranked in the Google suggestion list. Here is in Fig. 5, the clean and clean comparison shows that with and without translation with respect positive and total comments ratio with percentile ranking.

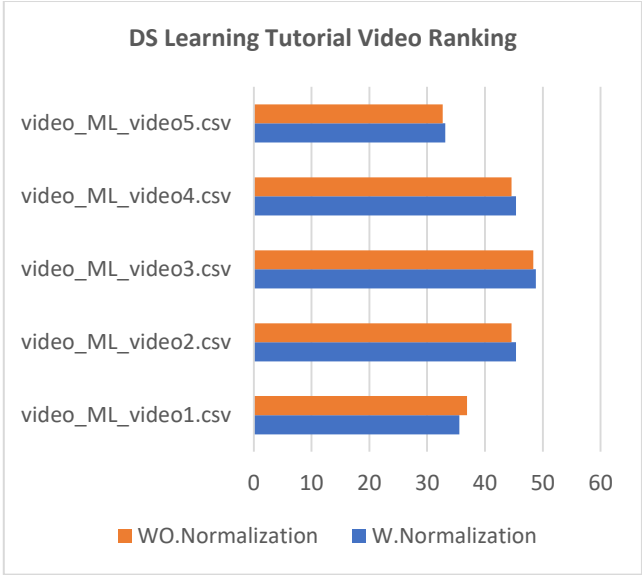

FIGURE 5: DS learning tutorial video ranking

Here is another tutorial about machine learning keyword here is the best one is video_2 is better than all and with translation we process the nomination of the values with on the basic of positive comments is shown in Fig. 6.

$$N_r = (P / T) * 100) \quad (1)$$

The second category is also showing that the method is effective and efficient and the comparison graph shown in Fig. 7.0 in which the Normalization is the score obtained by the equation (1). Video ranking in this scenario shows that the even the comments are in or other languages this method shows that we can get help from the methodology and produce the productive and qualitative video content.

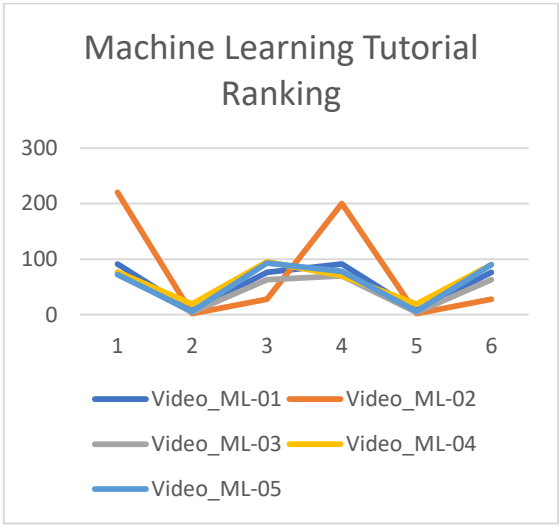

FIGURE 6: Machine learning

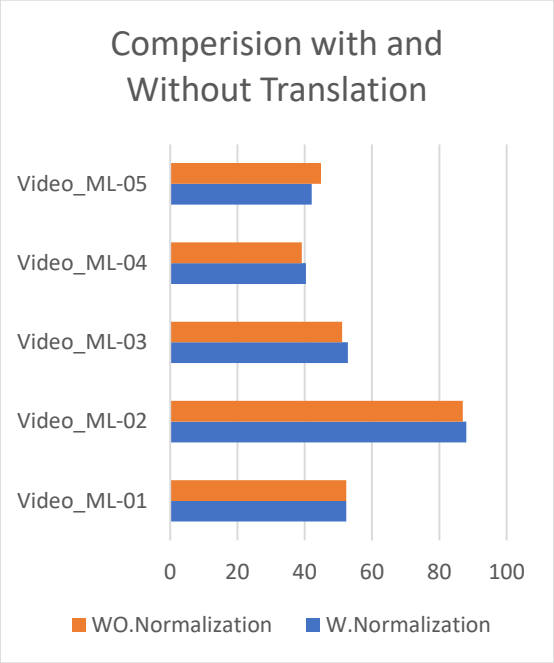

FIGURE 7: With and without translation.

V. CONCLUSION

The conversion of the comments and currently available packages are at some extend are productive and through the conversion we can find the more accurate feedback from the audience and this will help the user as well as the content producer to produce the more productive and qualitative videos. Our method is productive as like human work and results shows the method is so productive for user and content producer too.

REFERENCES:

[1] Abbas, Syed Manzar. 2017. "Improved Context-Aware YouTube Recommender System with User Feedback Analysis." *Bahria University Journal of Information & Communication Technologies (BUJICT)* 10 (2).

[2] Abu-El-Haija, Sami, Nisarg Kothari, Joonseok Lee, Paul Natsev, George Toderici, Balakrishnan Varadarajan, and Sudheendra Vijayanarasimhan. 2016. "Youtube-8m: A Large-Scale Video Classification Benchmark." *ArXiv Preprint ArXiv:1609.08675*.

[3] AMIT, CHAUDHARY. 2019. "NLP to SQL: TRANSLATING NATURAL LANGUAGE TO DATABASE QUERY LANGUAGE." PhD Thesis.

[4] Beghele, Sunilkumar N, and Pallavi V Kulkarni. 2019. "Using Deep Learning Technique To Query Relational Data Using Multi-Lingual Query Generator And Translator With NLP Support" 8 (08): 6.

[5] Chen, Yen-Liang, Chia-Ling Chang, and Chin-Sheng Yeh. 2017. "Emotion Classification of YouTube Videos." *Decision Support Systems* 101 (September): 40–50. https://doi.org/10.1016/j.dss.2017.05.014.

[6] Covington, Paul, Jay Adams, and Emre Sargin. 2016. "Deep Neural Networks for Youtube Recommendations." In *Proceedings of the 10th ACM Conference on Recommender Systems*, 191–198.

[7] Ezpeleta, Enaitz, Mikel Iturbe, Inaki Garitano, Inaki Velez de Mendizabal, and Urko Zurutuza. 2018. "A Mood Analysis on Youtube Comments and a Method for Improved Social Spam Detection." In *International Conference on Hybrid Artificial Intelligence Systems*, 514–525. Springer.

[8] "Google Trans Python Free API." n.d. https://pypi.org/project/googletrans/.

[9] Nawaz, S., M. Rizwan, and M. Rafiq. 2019. "RECOMMENDATION OF EFFECTIVENESS OF YOUTUBE VIDEO CONTENTS BY QUALITATIVE SENTIMENT ANALYSIS OF ITS COMMENTS AND REPLIES." *Pakistan Journal of Science* 71 (4 Suppl): 91–97.

[10] Nima, Prateek. 2020. "Automatic Filtration of Misleading Youtube Videos Using Data Mining Techniques."

[11] Papadamou, Kostantinos, Antonis Papasavva, Savvas Zannettou, Jeremy Blackburn, Nicolas Kourtellis, Ilias Leontiadis, Gianluca Stringhini, and Michael Sirivianos. 2020. "Disturbed YouTube for Kids: Characterizing and Detecting Inappropriate Videos Targeting Young Children." In *Proceedings of the International AAAI Conference on Web and Social Media*, 14:522–533.

[12] Poché, Elizabeth, Nishant Jha, Grant Williams, Jazmine Staten, Miles Vesper, and Anas Mahmoud. 2017a. "Analyzing User Comments on YouTube Coding Tutorial Videos." In *2017 IEEE/ACM 25th International Conference on Program Comprehension (ICPC)*, 196–206. IEEE.

[13] ———. 2017b. "Analyzing User Comments on YouTube Coding Tutorial Videos." In *Proceedings of the 25th International Conference on Program Comprehension*, 196–206. IEEE Press.

[14] "TextBlob: Simplified Text Processing." 2020. 2020. https://textblob.readthedocs.io/en/dev/.

[15] Tripto, Nafis Irtiza, and Mohammed Eunus Ali. 2018. "Detecting Multilabel Sentiment and Emotions from Bangla YouTube Comments." In *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*, 1–6. IEEE.

[16] "Valence Aware Dictionary and SEntiment Reasoner." 2020. *Valence Aware Dictionary and SEntiment Reasoner* (blog). 2020. https://pypi.org/project/vaderSentiment/.

[17] "YouTube for Press." 2020. *YouTube by the Numbers*. https://www.youtube.com/about/press/.