Pakistan Journal of Engineering and Technology, PakJET
Multidisciplinary | Peer Reviewed | Open Access
Volume: SI, Number: 01, Pages: 201- 204, Year: 2020

Qualitative Analysis of the YouTube Video in the Domain of Skin Care

Hafiz Naveed Hassan, M. Abubakar siddique, Safdar Hussain, M. Imran asad, and Jam Munawar Gul

Khwaja Freed University of Engineering And Information Technology, Rahim Yar Khan, Pakistan.
Corresponding author: Hafiz Naveed Hassan (e-mail: naveed.blouch99@gmail.com).

Abstract-YouTube ranking are the most impressive topic for researchers and the researcher try to get the qualitative content of the YouTube video without watching the video if the user is new its most important for us to enable the user to find the best video as per quality of the content for this purpose we proposed as model which gives us the classification of the comments with its class and also predict the how much the fake comments are written in the comments section to help the video content to ranked on the YouTube. This qualitative work helps the research after the preprocessing the comments and filter the fake comments and then find the qualitative comments from the video feedback section. Our proposed model help the user to classify his comments about the skin care videos the skin care means skin care and treatment and if there is some fake comments and fake ideas are putted in the video or comments this will classify and help the user about the video with pure qualitative analysis of the video with comments and this helps us to truly rank the video and the user will get benefits truly. With the help of Machine learning algorithms and feature engineering we meet the target and after this we received the 0.94 F1 score.
Keywords: *fake comments, YouTube, Machine Learning algorithm, multi-class classification*

I. INTRODUCTION

YouTube is the platform when any user needs to learning or need some information and want to see something and he search on the YouTube. The huge content of this platform helps the user to get the exact same content but the YouTube for the help and support shows the suggested or related videos based on geo location and user likeness history [1]. YouTube is the one platform that have two billion active users and more than 70% user are come from mobile devices 15% of the traffic of the YouTube are come from USA [2]. It is accepted that 12mints per day averagely each user spends the time on YouTube. It is noticed that the 70% people are watch the video as suggested by Google search engine. Averagely the YouTube user visit the 5-6 page daily in 2020 year the YouTube earn $5.5 billion from advertisements. YouTube is the second largest TV watching application in the age from 18 to 34 [3]. Today the YouTube is most widely used online social media platform shown in Fig. 1.

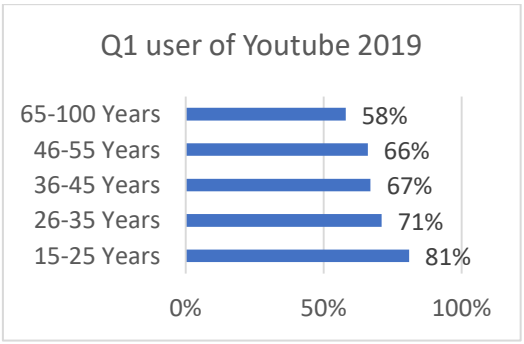

FIGURE 1: YouTube users

If we consider the country wise usage then the 15% are from USA and 8.1% from India and 4.6% from Japan and the 3% video watcher in the India are watching the YouTube videos. Here are the statistics shown in Fig. 2.

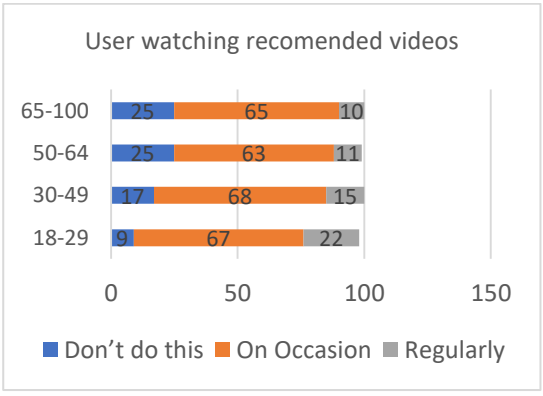

FIGURE 2: YouTube statistics

Pew research finds that the 79% views are from the top 10 most popular videos and the 21% are remaining for the other non-popular videos [4]. From the statista.com the reported the only skincare viewer on the YouTube from 2009 to 2018 is shown in Fig. 3.

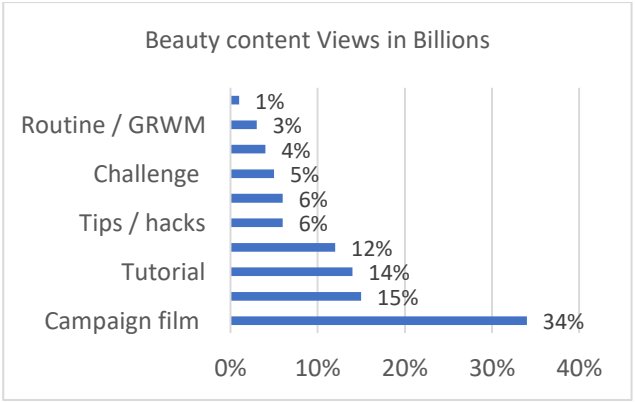

FIGURE 3: Skincare viewer on YouTube

We see here the beauty content watch is so productive in coming days and also the beauty care products business also growing up. We notice that the 14% on the YouTube is the people who are watching the YouTube due to learning and our focus is the basically on this when the people use the YouTube and try to learn something but the user how to know the best practical experienced tutorial is on the YouTube.

A) Our Participation in Skin Care domain:

1. Find the productive comments
2. Classify the comments
3. Classify the fake comments

II. LITERATURE REVIEW

Comments classification is the done with the help of machine learning algorithms. The Machine learning algorithms are classifying the comments based on the trained and test data the trained data used for training the machine learning model in the study working on the fake comments classification using the 2 algorithms one is Naïve Bayes and other one is Logistic regression using the two very popular data GUI tools Weka tool and Rapid miner and the best are received from the Weka the Naïve Bayes having accuracy 87.21% and the logistic regression is 85.29% and in the Rapid miner the result are changed and go down at NB is 80.88% and LR is 80.41% [5].

We when required a training data we should need to collect or scrap for YouTube and then perform annotation for Fake 0 and original comment is 1 used as encoding processes for batter and fast machine earning. Another study helps the user to classify the comments in the domain of cooking recipes In this study, they develop a real-time system to extract and classify the YouTube cooking recipes reviews automatically [6]. This system is based on Support vector machine approach and deals with the social media text characteristics. The proposed system collects data in real time from YouTube according to a user request. To improve the performance of our system we proposed some algorithms that constructed on sentiment bags, based on emoticons and injections. And their model accuracy was 83.5% [7].

People need to know the informative comments and they need to know the what the topic is discussed. The discussion is in the context for outside the context the study helps the user to find the informative comments they offer the novel approach in which 20 sampled videos are selected from TED Talk 1861 videos with 1000 sample comments and originally dataset consist on the 380619 comments [8-10].

YouTube also have some aspects which are like, dislike, number of views and watch time these aspects or features can be used to predict the relevancy of the video or irrelevancy of the video the study shows this is implemented and the results shows that the F1 score 0.74 [11-13].People use the Java library which is mallet library used to extract the useful comment related to the video title and they remove the non-English comments and the Neural network predict the correct classification with 87.46% with the weight and topic extraction of the comments(Chauhan and Meena 2019). In the studies we faced that the video recommendation system are can works on the comments based and new video may or may be included for recommendation and third one is these model cannot work with large scale videos and so the Deep Neural Networks helps us to recommend the video related to qualitative recommendation they used the many full connected rectified linear unit which gives us efficient and effective results [3,5].

III. METHODOLOGY AND DEPLOYMENT

Here is our methodology the we are adopted for out proposed work. We used the data scrapping the tools which are Python package so called Beautiful-Soup this package and Google Chrome data scrapping tools which is we are used is instant scrapper both are very useful for us to collect the comments shown in Fig. 4.

	comment	class_pred
0	cleansing	1
1	okay buy ingredient mask	3
2	hi got open acne scar quit dry	4
3	mam large due pimple solution helpful meplz reply	6
4	remedy daily	3
5	acne di please tell treat	6
6	made glowup lockdown	1
7	sensitive also	5
8	look beautiful without makeup	1
9	thank u much dear clear problem actually dont ...	1

Fig 4: Comment from scrapping tool.

After collection we clean the data and after preprocessing the data, we use the classes to annotate the data the classes are listed below.

1. Beauty feedback Positive
2. Beauty feedback Negative
3. Treatment Feedback Positive
4. Treatment feedback Negative
5. Warning / Danger feedback
6. Questions as Feedback
7. Fake feedback
8. Non-Feedback

These classes are annotating the data and use the label encoding to encode the classes and then the dataset prepared for machine learning after creating the features. For the features engineering we use the two feature engineering extraction methods [13].

1.Conunt Vectorization

2.TF-IDF Vectorization

We use the 80% for model training purpose and the 20% for model test purpose. We use the multi-classes to prediction we use the Logistic regression and random forest for class prediction and we deploy the combined and separate for both category one for skin care classes and 2nd one is fake and non-fake comments.

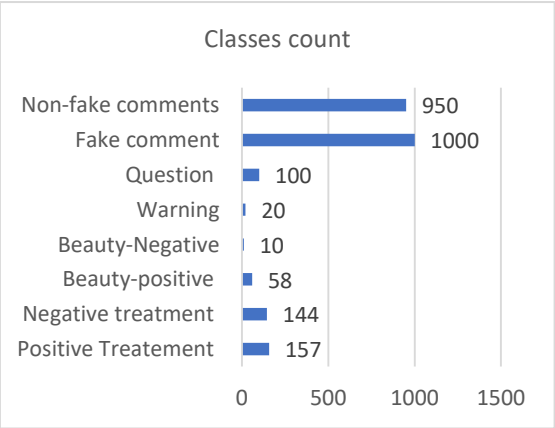

FIGURE 5: Classes count

Meaning of the classes are non-fake means this comment is not fake and fake means this is fake comment and written by the content creator to make ranking powerful based on the comments shown in Fig. 5 and the Beauty feedback positive means that the comment is about the beauty care and the user use this tip and guide and based on his personal experience he get positive response and the comment on the YouTube for users to guide them the method or tip described in the video is so effective and the same this the Negative feedback about beauty care tip this tip or method is experienced but the user cannot get the good benefits from the from the tip or method and the warning means the user don't use or do not rely on the video author talks and ideas and don't try this question is the class which means the people watch the video and they need for help to implement this as personal experience. The processes model and flow are shows in Fig. 6.

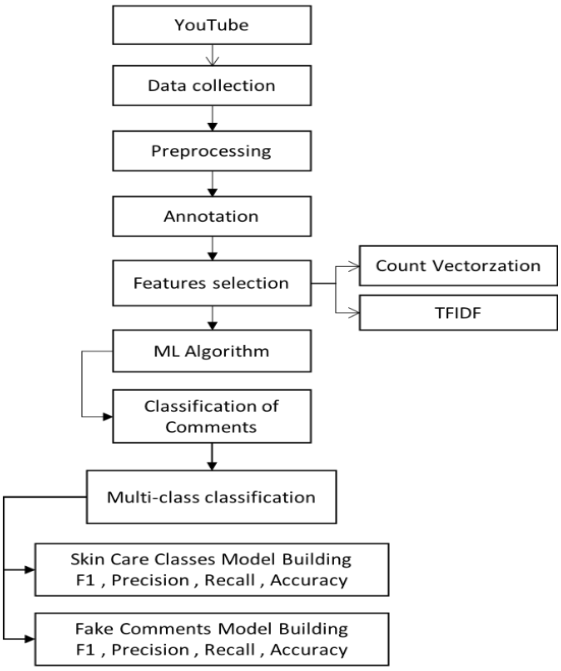

FIGURE 6: Process model

Processes model helps us to describe the methodology in pictorial view.

IV. RESULTS AND DISCUSSION

We deploy the different machine learning algorithms the below results are shown in Fig. 7. When we see the results, we notice that the fake comments are better classify but the other classes are not good same as fake comments because our classes are not balanced and that why the classification results are not same as fake comments. We notice that the TF-IDF is productive for us we also use the count vectorization and word2vec but the results are below then 50% so wo excluded the results shown in Fig. 7.

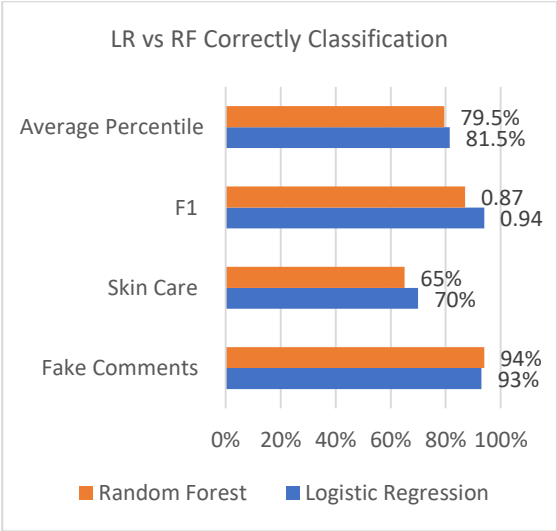

Fig 7: LR vs RF correctly classification

V. CONCLUSION

Our proposed method help the user to find the best video experienced by persons and stop use the products and tips that are given in the video content we also find the fake comments if someone need to increase the subscriber and increase his views and watch time the new user will be safe from this blunder and the original video and its contend will be reached with the actually demanding customer and user. Model results shows that the methodology is so productive and so efficient same as like human woks.

REFERENCES

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

[2] "YouTube Statistics." n.d. *Source: Cowen Proprietary Tracking Survey , n =~2500,Feb 2019.Other Includes Social Networks and Other Video Platforms*. (blog).

[3] Bärtl, Mathias. 2018. "YouTube Channels, Uploads and Views: A Statistical Analysis of the Past 10 Years." *Convergence* 24 (1): 16–32.

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

[5] Aziz, Aqliima, Cik Feresa Mohd Foozy, Palaniappan Shamala, and Zurinah Suradi. 2017. "YouTube Spam Comment Detection Using Support Vector Machine and K–Nearest Neighbor." *Indonesian Journal of Electrical Engineering and Computer Science* 5 (3): 401–408

[6] Benkhelifa, Randa, and Fatima Zohra Laallam. 2018a. "Opinion Extraction and Classification of Real-Time Youtube Cooking Recipes Comments." In *International Conference on Advanced Machine Learning Technologies and Applications*, 395–404. Springer.

[7] ———. 2018b. "Opinion Extraction and Classification of Real-Time Youtube Cooking Recipes Comments." In *International Conference on Advanced Machine Learning Technologies and Applications*, 395–404. Springer.

[8] 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*.

[9] Bessi, Alessandro, Fabiana Zollo, Michela Del Vicario, Michelangelo Puliga, Antonio Scala, Guido Caldarelli, Brian Uzzi, and Walter Quattrociocchi. 2016. "Users Polarization on Facebook and Youtube." *PloS One* 11 (8): e0159641.

[10] Chauhan, Ganpat Singh, and Yogesh Kumar Meena. 2019. "YouTube Video Ranking by Aspect-Based Sentiment Analysis on User Feedback." In *Soft Computing and Signal Processing*, edited by Jiacun Wang, G. Ram Mohana Reddy, V. Kamakshi Prasad, and V. Sivakumar Reddy, 900:63–71. Singapore: Springer Singapore. https://doi.org/10.1007/978-981-13-3600-3_6.

[11] Choi, Seungwoo, and Aviv Segev. 2020. "Finding Informative Comments for Video Viewing." *SN Computer Science* 1 (1): 47.

[12] 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.

[13] Uysal, Alper Kürşat. 2018. "Feature Selection for Comment Spam Filtering on YouTube." *Data Science and Applications* 1 (1): 4–8.