

Monetisation Of Youtube Content Using Data Mining Techniques

Kanika Singhal, Abhineet Anand

Abstract: The race of online business is increasing day by day. No one wants to sit ideal at home, people just wants to do something to earn money. Online entrepreneur is one such thing. Starting YouTube channel can be one such destination. Once a YouTube channel is created and it meets the prescribed guidelines decided by YouTube, the channel is reviewed and if it meets the programs and policies of YouTube, the monetisation is enabled for the channel. After the monetisation on the channel is enabled, it can be linked to AdSense account and the revenue can be generated. This paper provides review of various data mining techniques used in the monetisation of YouTube content. In this analysis, we also describe some methods to improve the performance of monetisation by using data mining techniques

Index Terms Data mining, YouTube, Monetisation, Association, Regression, Clustering, Prediction.

1. INTRODUCTION

Monetisation means you allow YouTube to run advertisements on your posted videos and the advertisers are charged to run these advertisements on the videos on your channel, the charges are shared between the channel owner and YouTube. People are becoming more and more dependent on YouTube nowadays. Be it cooking, dancing, reviews of products, parenting, and travelling vlogs, anything that can come into one's mind can be found on YouTube. People learn a lot from Youtube and the good videos which people relates to goes viral within a second. So starting a YouTube channel can be a good idea and if you achieve success and become a famous youtuber, your channel content will be monetised and you will start making money. Definitely, it is not an easy process but consistency and determination can lead to good results. Monetisation of the content is not that easy as it seems to. Diverting traffic to your videos requires tremendous efforts. YouTube success is not an overnight journey. A lot of efforts is required for the monetisation of the content. First of all high quality content is must. Constant uploading of videos and communication with the viewers are the base of the channel. Starting from shooting the video, editing, video processing, optimising the content and posting it on your channel refines your YouTube skills [1]. The videos should be valuable to the viewers watching your YouTube channel. Passionate and concise content is a must thing in developing your YouTube channel. it can only lead to monetisation of the content and earning revenue from the videos. Checking your channel analytics regularly is also a must do thing. The YouTube provide these analytics on your YouTube channel page. This includes watch time of your videos, views on your videos, the subscriber gain on your channel and the estimated revenue generated from your videos. The watch time of the video depends on the people watching video .If the video becomes famous, the watch time will surely increase. The views on the video depend on how many people relates to that video. If the video relates to more people, the views will considerably increase with time. The subscriber gain depends on the people

watching your video and if the content relates to them, they will surely subscribe the channel. The estimated revenue of the channel depends on the advertisements running on the channel. The advertisements will be generating revenue for the channel. Figure 1 shows the analytics of a YouTube channel.

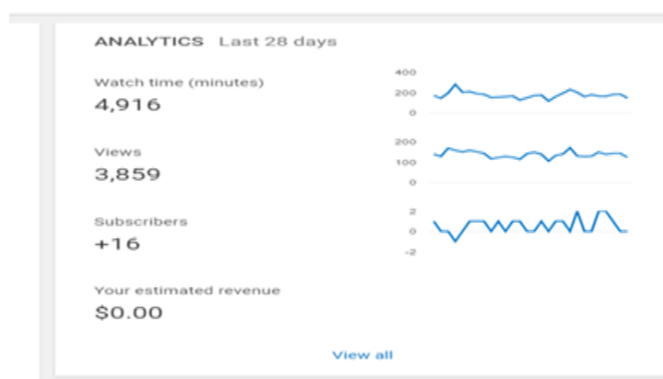


Figure 1: Analytics of a YouTube channel

Data mining in online media content is current research topic because online data is increasing day by day with a very high pace. There is a diverse community of users and the web is too dynamic. Some of the live examples that we can remember are the Kerala floods recently. A huge amount of money was collected online through social media. Such is the power of online media. So there is a need to extract useful data from this data to make them available to the user by applying various associations, correlations, clustering, partitioning, classification and prediction. Data is the new currency nowadays [1]. Data mining retrieves useful information from the data sets high in volume. The mining of correlations, clusters, and frequent patterns can also be observed by data mining. Data mining techniques can help improve the monetisation process. The techniques such as clustering, regression, outlier detection, association rules can be applied to monetisation. These techniques are applied to the monetisation process to generate better results. The detailed analytics includes the watch time, average view duration, views, and estimated revenue, likes, dislikes, comments, shares on your videos, the videos in the playlist and the subscriber gain. Figure 2 shows the detailed analytics.

- Kanika Singhal is currently pursuing Doctoral degree program in Computer Science and Engineering in Galgotias University, Greater Noida, UP, India. E-mail: kanikasinghal28@gmail.com
- Abhineet Anand is currently working as Professor in Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India. E-mail: abhineet.anand@chitkara.edu.in, abhineet.mnnit@gmail.com

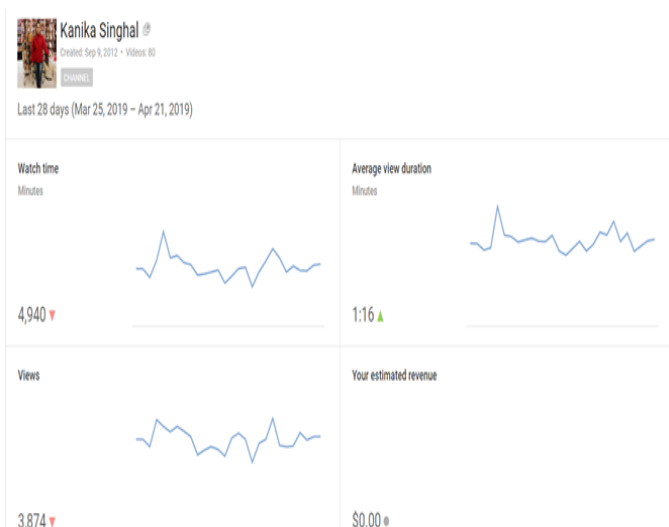


Figure 2: Detailed analytics of a YouTube channel

2 LITERATURE REVIEW

The first step that is necessary to continue this study is the study of previous technologies that exist in the system. This is the building block of any study that has to be continued. The process of literature review is not the raw form of data. Infact it is the study of the existing approaches. It has to be in the selective and summarized form. To explore the data mining techniques in online media content, firstly we have to study the techniques that already exist in the hierarchy. The motive behind this work is to explore data mining techniques that are suitable for solving complex online data problems and to make this technology available to the user. For this purpose an intensive study of existing approaches is required. So the various data mining techniques studied for this work are described below with their features.

3 TECHNIQUES USED IN DATA MINING

As demonstrated in this document, the numbering for sections upper case Arabic numerals, then upper case Arabic numerals, separated by periods. Initial paragraphs after the section title are not indented. Only the initial, introductory paragraph has a drop cap.

3.1 PREDICTION

Prediction predict continuous valued functions. For example if a student has taken admission in class tenth of the school then he will be getting pass or fail in board examination is a prediction or a person doing good in academics at some stage of life will be successful or not in his coming life. In this case a predictor will be constructed that will predict the continuous valued function. This techniques is also used in data analysis. There are different issues that exist in the prediction methodology. Some of them are data cleaning, data reduction and relevance analysis. Data cleaning removes the noise and treat the missing values with the most probable or unknown value. Data reduction consists of normalisation and generalization. In relevance analysis, the irrelevant data is removed from the database[1].These are the issues which are addressed during prediction approach.

3.2 CLUSTERING

Cluster is the object of the same class. While doing cluster

analysis, we first partition the set of data into groups based on similarity of the data and then assign the points to different clusters [4]. The main advantage of clustering is that the changes can be incorporated very easily. Fig 1 comprise of circles with have oval shaped points inside it. These points belong to the same class so they are contained in one cluster. The point lying outside is known as outliers. Cluster 1 contains same class objects. Cluster 2 contains a single object. The point that does not belong to any cluster is outlier. The outlier analysis is done by block and consecutive methods. They form the multidimensional array and the points outside the boundary are considered. Fig 1 is the proper illustration of clustering process⁴.

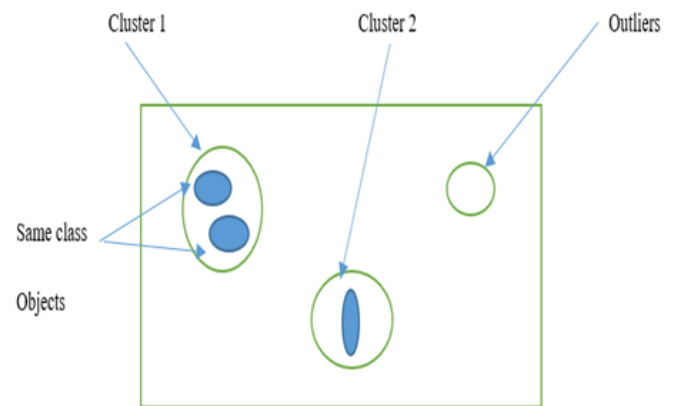


Figure 3: Cluster and Outlier

3.3 BINING

It is based on the concept of neighbourhood. The values are grouped together into buckets or bins. The creation of buckets is based on equal frequency method [3]. Each value of the bin is replaced by the closest value be consulting its neighbourhood. For example if we consider the value set

$$A = \{a, b, c, d, e, f, g\}$$

Some values in this set are on equal frequency and some are not. Let us consider a, d, f, g belongs to one group and the values left belong to other group. So there will be two groups formed from this single set.

$$A_1 = \{a, d, f, g\}$$

$$A_2 = \{b, c, e\}$$

So, Binning helps to remove background noise and partition the data into different sets.

3.4 REGRESSION

The data can be fitted in the cluster using regression function. The process for estimating the relationships among variables is known as regression analysis [6]. The data set is estimated carefully and then regression is applied. In data mining, regression can be used for both classification and prediction. The difference is that regression in classification can be used to define categorical value and regression in prediction is also used to predict the continuous values[4].Regression can be of two types that is linear and multiple regression. Linear regression is dependent on one variable and its graph gives a straight line while multiple regression is dependent on more than one variable.

3.5 AGGREGATION

The data from various sources are collected and aggregated into a common source. The source of data can be homogenous or heterogeneous depending on the

requirement. The data is then collected and aggregation function is applied on the data is gathered and summarised in report format and the data is make presentable to present it to the stakeholder. Aggregation simply means how the data is gathered and then analysed summarised and presented to the user. In data base technology many functions can be applied on it such as count, average, sum, minimum and maximum. But in data mining the techniques are different.

3.6 NORMALISATION

This technique allows different parameters to fall into a specific range. For example the binary values lies between 0 to 1. The range may vary from approach to approach. A limit is specified with in the approach. It rescales the value into a particular range. Adjusting the value in the range described is the priority of normalisation. In other words, we can say, normalisation helps to remove the redundant data from the database [7]. As redundant is a matter of great concern in any field of data, it is necessary to remove it. This can be done by normalisation to a great extent.

4. MONETISATION PROCESS

Monetisation is enabled on the YouTube channel when the channel has at least 1000 subscribers and 4000 watch hours in previous 12 months [2]. After meeting this criterion, the channel goes for review and after being reviewed according to the program and policies of YouTube, gets monetised. After monetisation of the content, the advertisers will put advertisements on your videos depending on the type of video posted for example if you have posted a cooking recipe video, then it might be possible that the advertiser put the advertisement of spices in your video. The earning of the revenue from the videos not much depends on the views and subscribers but on the watch time of the advertisements placed in between the video. Advertisers only pay when the advertisement on your video is watched for at least 30 seconds [2]. After monetisation is enabled on your channel, your channel should be connected to AdSense account to earn revenue. Figure 3 shows the step by step process of monetisation.

5. SETTING UP AN ADSENSE ACCOUNT FOR MONETISATION

Once the monetisation is enabled for the channel, it's time to earn revenue through the advertisements that has been placed in the videos. To do this, it is necessary to set up an AdSense account.

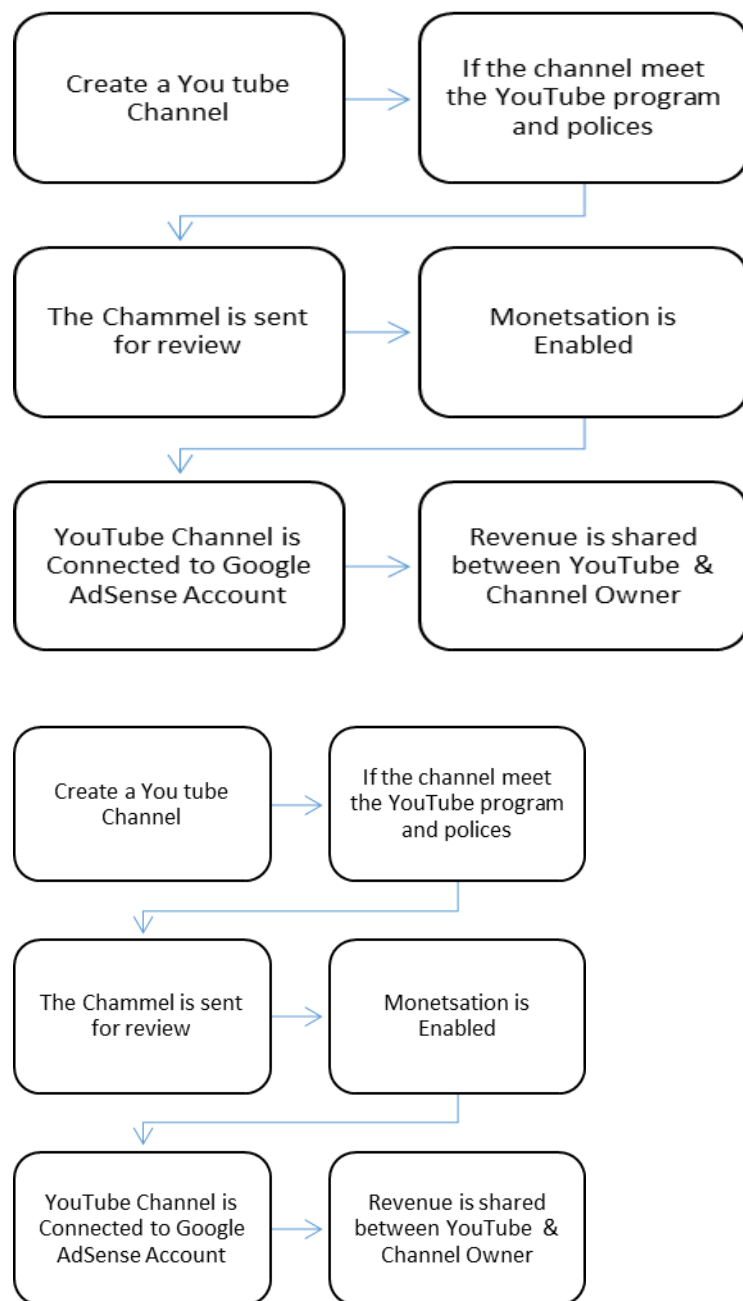


Figure 4: Step by Step Monetisation

The Google account by which you have made your YouTube channel is to be linked with AdSense [3]. You have to show the type of content which you will be displayed on your YouTube videos. Then you have to fill the application form for linking your YouTube channel to AdSense account and submit the application. The AdSense application will be reviewed and you will receive an email once it is reviewed. Once it is approved, the YouTube channel is linked to AdSense account and the channel start generating revenue. Figure 4 shows the step by step procedure of linking YouTube channel to AdSense account.

6. DATA MINING APPLICATIONS IN MONETISATION

The main concern of monetisation of the content on YouTube

is creating high quality content videos that attracts a large number of subscribers and more views on the videos that can lead to generation of income from these videos. This can be easily possible if we follow some steps and keep something in mind

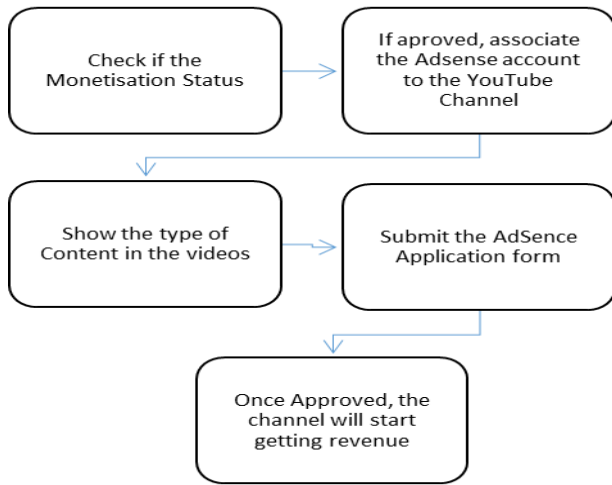


Figure 5 Step by step monetisation process

6.1 Shooting the video

The foremost thing about a video is shooting it in a proper manner. The video must be shot with a good camera and it should not be shaky or of poor quality. This can be avoided by using the shot boundary detection technique of data mining where the video is segmented in between the shots of the camera to detect the boundaries and to produce high quality videos [4].

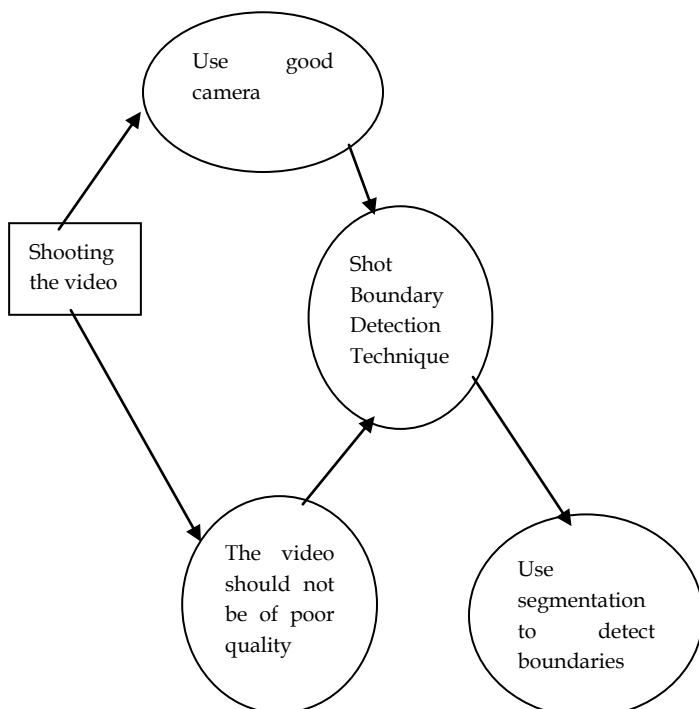


Figure 6: Shooting the video using shot boundary detection

1.1 Video processing

It is the most important step after shooting the video. The long frame video must be segmented and the data must be divided into small clusters and then sub clusters. The processing on the data will become easy if the long frame videos are divided into small time frames for editing using clustering approach of data mining. The implementation of operation becomes easy if the videos are divided into small time frames. The unwanted portions can be omitted and the data can be processed very easily.

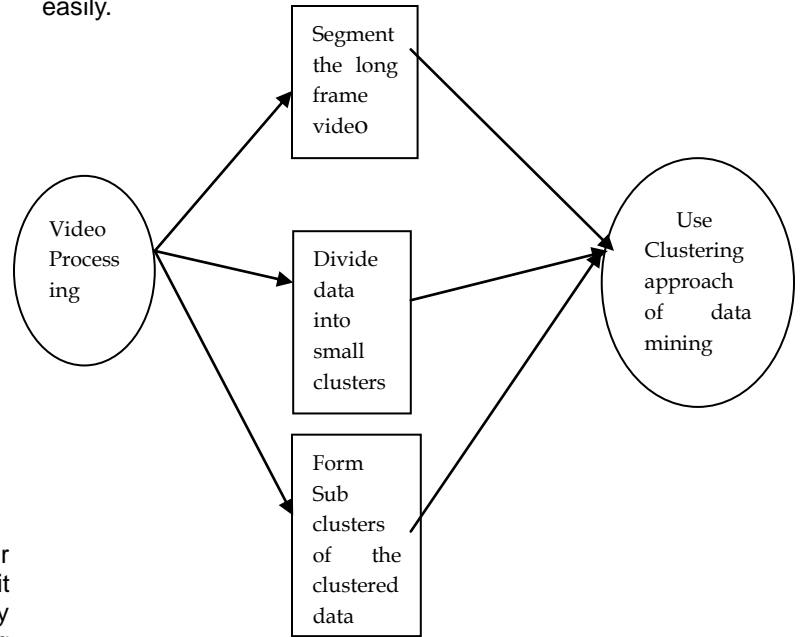


Figure 7: Video processing using clustering

1.2 Video Information retrieval

After processing the video, not every information on the video is relevant. Suppose if we want to omit certain portion of the video that is not in relevance and the relevant information needs to be retrieved [5]. This can be done by dividing the video into smaller units and forming a hierarchy using hierarchical clustering method of data mining. Forming a hierarchy and then retrieving the necessary information will be a simple task.

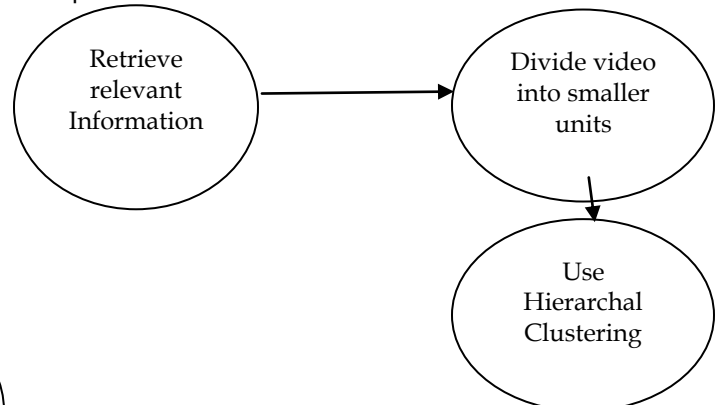


Figure 8: Video information retrieval using hierarchal clustering

1.3 Video Summarization

This refers to the concise content that we create for our target audience. The viewers should understand the content in the

proper manner and it should seem useful to them. Here rule based classification can be used to extract information from the videos that draw the attention of the viewer viewing it.

2. WORK ANALYSIS

Data mining is the extraction of information from large data sets [6]. Various data mining techniques are used in monetisation. Table 1 shows the data mining techniques, monetisation current status in relation to these techniques and the scope of improvement in the monetisation process. Classifications classify the content of the video in accordance with the policies of YouTube and it can be improved better if we can track the pattern of the content appearing on the YouTube channel. The tracking of pattern will be helpful in determining the viewer's interest in different types of videos.

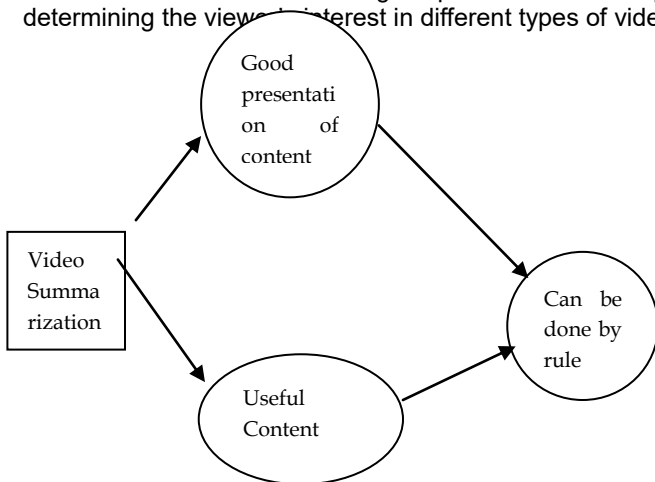


Figure 9: Video summarization using rule based classification

Clustering forms the cluster of the viewers who has the same interest for example some subscribers may be interested in cooking videos so they can be put into one cluster, other subscribers may be interested in travel videos so they can be put into another cluster. Cluster will group the viewers according to their interest. The process can be made better if the data is normalised [6]. Regression helps to track the pattern of the viewers watching the content and then they become the subscriber if they find the content appealing. The performance can be optimised by removing the outliers and focussing on the subscribers. Outer detection refers to removing the content that is not authorised. The content that is not in accordance with YouTube programs and policies is rejected [7]. If the unauthorised content is deleted beforehand before posting it on the channel, the performance will be greatly improved. Sequential pattern finds the relevant pattern that governs the posted content on the channel. For example observing the types of videos coming on the channel, the days when the video is posted. If the time and space overhead is decreased, it will optimise the data to a great extent. Predict predicts the number of subscribers on the channel, analytics; watch time of the videos, revenue generated from the videos and many other factors. If we can predict the performance improvement factors of the channel, it may lead to the improvement of the channel further. Association rules analyse the rules under which the video is shoot, the conditions of editing and the optimisation of video [8]. If we can control the preconditions before shooting the video and post conditions after shooting the video, quality content can be generated.

Table 1: Data mining techniques to improve monetisation process

Techniques	Monetisation Current Status	Improvement in Monetisation process
Classification	Classify the content of the YouTube Channel Videos in accordance with YouTube program and Policies.	The content can be better classified by tracking the pattern of the content appearing on the channel
Clustering	Forms cluster of the viewer's viewing the channel as per their interest.	The cluster can be better formed if the process of normalization is followed.
Regression	It helps to track the pattern of how many viewers visiting the channel become its subscribers.	The performance can be optimized by removing the outliers and focusing on the subscribers.
Outer detection	The content to be monetized should be in accordance with YouTube program and policies. The content not according to that is rejected.	The mechanism can be improved by removing the unwanted content beforehand.
Sequential Pattern	Finding relevant patters occurring in the sequence can be found in the content posted on the channel.	The Time and space overhead can be decreased using optimal version of sequential pattern.
Prediction	Predictions regarding the subscribers of the channel, viewers, watch time and generated revenue.	Prediction of performance improvement can be made which can lead to a better content.
Association Rules	The rules that govern the content of the videos such as the conditions in shooting the videos, editing it and posting on the channel.	The content can be greatly improved if we know the preconditions and post conditions.

CONCLUSION

Data mining technologies are greatly applied to content monetisation but the scope of improvement still remains. There are some areas that need to be addressed here. Firstly, the creator of the YouTube channel has no control over the advertisements that are placed on the monetised content and the creator cannot control the watch time of the advertisements [9]. It completely depends on the viewer. Secondly, the earning from the advertisements is unpredictable. There is no metric as to tell how much commission will be earned through these advertisements. Thirdly, it is difficult to predict the type of content the audience would like to watch, here increasing subscribers and watch hours becomes a great challenge.

REFERENCES

- [1] Anwar F, Petrounias I, Morris T, Kodogiannis V (2012) Mining anomalous events against frequent sequences in surveillance videos from commercial environments. *Exp Syst Appl* 39:4511– 4531
- [2] Anjulana A, Canagarajah N (2009) A unified framework for object retrieval and mining. *IEEE Trans Circ Syst Video Technol* 19(1):63–76
- [3] Ahmed A (2009) Video representation and processing for multimedia data mining. *Semantic mining technologies for multimedia databases*. IGI Press, pp 1–31

- [4] Kumar, V. and Sisodia, A. and Maini, U. and Pankaj and Anand, A.(2016) "Comparing algorithms of community structure in networks", Indian Journal of Science and Technology, 9, 44.
- [5] Anjulan A, Canagarajah N (2007) A novel video mining system. In: Proceedings of 14th IEEE international conference on image processing, San Antonio, Texas, pp 185–189
- [6] Aradhye H, Toderici G, Yagnik J (2009) Video2Text: learning to annotate video content. In: Proceedings of IEEE international conference on data mining workshops, pp 144–152
- [7] Bhatt CA, Kankanhalli MS (2011) Multimedia data mining: state of the art and challenges. Multimedia Tools Appl 51:35–76
- [8] KUMAR, Vikas et al. Comparing Algorithms of Community Structure in Networks. Indian Journal of Science and Technology, [S.I.], nov. 2016. ISSN 0974 -5645. Available at:<<<http://www.indjst.org/index.php/indjst/article/view/105088>>>. Date accessed: 27 Feb. 2019. doi:10.17485/ijst/2016/v9i44/105088.
- [9] V. K. Sihag et al., "Detecting community structure based on traffic at node in networks," 2014 IEEE Students' Conference on Electrical, Electronics and Computer Science, Bhopal, 2014, pp. 1-9. doi: 10.1109/SCEECS.2014.6804502.
- [10] Abhineet Anand, A. S. (2017). Data Leakage Detection Using Cloud Computing. International Journal of Engineering and Computer Science, 6(4). Retrieved from <http://www.ijecs.in/index.php/ijecs/article/view/3785>.