Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 198 (2022) 23–32

Procedia Computer Science

www.elsevier.com/locate/procedia

The 12th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2021)
November 1-4, 2021, Leuven, Belgium

# YouTube Monetization and Censorship by Proxy: A Machine Learning Prospective

Anthony Zappin[a], Haroon Malik[a]*, Elhadi M. Shakshuki[b], David A. Dampier[a]

[a]*College of Engineering and Computer Sciences, Marshall Univeristy, WV, USA*
[b]*Jodrey School of Computer Science, Acadia University, Wolfville, Canada*

**Abstract**

As consumption of digital content has climbed, so has censorship of the content. The censorship has only increased with companies more sensitive to the type of content that they tie their advertising to on digital platforms. Demonetization of videos is a primary way content is censored on YouTube. Demonetization, often referred to as "Apocalypse", is a process in which content creators, are denied paid ads in their YouTube videos. Consequently, they are denied revenue, their income on the video-hosting platform is reduced and their video is less likely to be promoted or recommended on the platform, eventually getting censored. YouTube's censorship algorithm is not public and is a Blackbox to the world. The paper proposes a methodology that employ four machine learners, i.e., C 4.5, Random Forest, Linear Regression and Support Vector Machine, to predict if changes in the meta-data of the YouTube video will lead to (demonetization) censorship of the video. Our methodology requires little time to train and achieve an accuracy of up to 87%. The methodology is useful to content creators trying to determine what content to create to maximize their revenue. As well is helpful to free speech advocates who may believe content is being unfairly or unlawfully censored.



*Keywords:* YouTube; Machine Learning; Random Forest; Liner Regression; Censorship; Demonetization

* Corresponding author. Tel.: +-304-696-5655.
*E-mail address:* malikh@marshall.edu


10.1016/j.procs.2021.12.207

## 1. Introduction

In today's digital age, content is increasingly being consumed online. Some of this content comes in the form of news, opinion, and entertainment. The largest online platform consumers turn to for content is YouTube. Indeed, YouTube has over 1 billion users log in monthly and over a billion hours of video are watched daily. As consumption of digital content (via the internet) has climbed, so has censorship of that content.

*Internet censorship* refers to a plethora of tools and strategies to prevent information from reaching users. It is the control or suppression of what can be accessed, published, or viewed on the internet enacted by regulators or on their own initiative. Internet censorship puts restrictions on what information can be placed on the internet or not. YouTube censors content for what it deems violations of the Community Guidelines using three primary methods. These methods are (i) outright content removal, (ii) channel removal through a three-strike system and (iii) demonetization.

Demonetization of videos is a primary way content is censored on YouTube. *Demonotization,* often referred to as "Adpocalypse", is a process in which content creators are denied paid ads in their YouTube videos. Consequently, (i) they're denied revenue, and their income on the video-hosting platform is reduced, and (b) their video is less likely to be promoted or recommended on the platform, eventually getting censored. YouTube's demonetization of content can come in two forms: (i) demonetization of an entire channel; or (ii) demonetization of a video-by-video basis. The paper details the YouTube Monetization and Censorship is detailed in Section 2.

This censorship has only increased with companies more sensitive to the type of content that they tie their advertising to on digital platforms. Escalating political divides in USA have also contributed to the rise in censorship on platforms such as YouTube. More specifically, political groups allege that YouTube (and other social media sites) not only have a political bias but targets and censor content from political ideologies.

### *1.1. Problem Statement*

We formulate the problem statement by listing some of the scrutinize and challenges the YouTube monetization process faces.

***Flawed Monetization Policy*** — Creators of the content on YouTube must abide by its community guidelines, terms of service, copyright policy, Google Ad-sense program policy, channel monetization policy, and ad-friendly content guidelines. While abiding by these policies may seem like the sure way of preserving the rights to video monetization, YouTube's monetization policy is flawed in its enforcement, i.e., YouTube can still limit the monetization or remove it on a video-by-video basis.

***Biased Algorithm*** — Around 4000 hours of video are uploaded to YouTube every minute; 65 years of video a day. To deal with the censorship of such a large amount of digital content on the platform, YouTube deploys an algorithm on the videos and its related meta-data, i.e., customized feeds, comments, and views to stream like recommendations and content moderation, i.e., censorship. The algorithm is collectively maintained by thousands of people who work for Google (software engineers, researchers, and content moderators) and millions who participate on the platform, create content, and help train the algorithm, which has not been public to date. By keeping the algorithm and its results under wraps, YouTube ensures that any patterns that indicate unintended biases or distortions associated with its algorithm are concealed from public view. By putting a wall around its data, YouTube, which is owned by Google, protects itself from scrutiny.

***Undisclosed Data*** — Though the censorship algorithm is not public; YouTube does not even share the results of the trained censorship algorithm, i.e., aggregate data revealing which YouTube videos are heavily promoted by the algorithm or how many views individual videos receive from "up next" suggestion. Disclosing that data would enable academic institutions, fact-checkers and regulators (as well as journalists) to assess the type of content YouTube is most likely to promote.

***Limited Research*** — Much of the research on YouTube content creators has focused on popular creators who earn a living on the platform. While there has been some study about YouTube's censorship policies and monetization policies separately, the use of demonetization of content as a proxy for censorship has largely been unexplored, particularly where political bias might play a role. There has been a significant amount of magazine articles and blogs that discuss social media platforms such as YouTube employing machine learning algorithms to

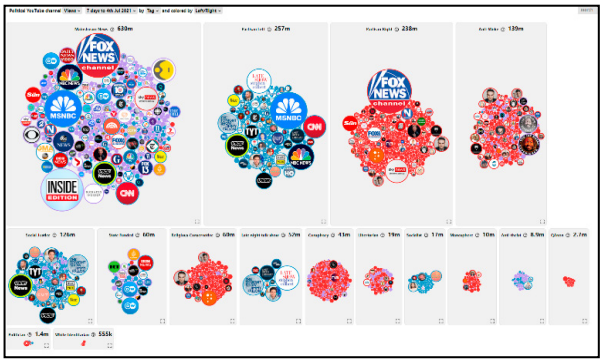


Fig. 1. Classification of videos via Transparency.Tube.

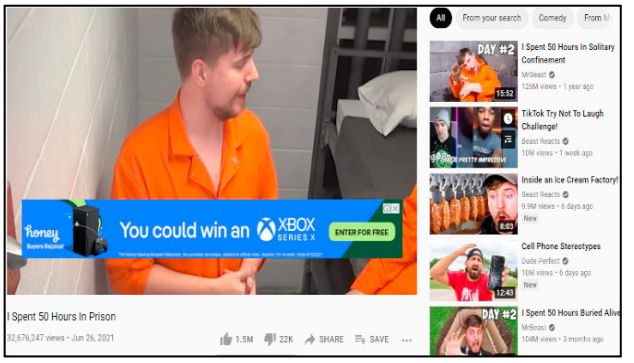


Fig. 2. Advertizement of Mr. Beast (monetized) video.

censor content through demonetization. However, there appears to be very little academic research on this topic. Given the importance of this topic and how it is at the forefront of social and news media, this is quite surprising.

### *1.2. Paper Contribution*

This paper employs machine learning techniques to gain insight into YouTube's demonetization algorithm, which has been alleged to be a proxy for censoring content. The paper aims to use machine learning techniques to determine whether political ideology plays a role in YouTube's monetization algorithm for individual videos. The results of the proposed work will be helpful for content creators trying to determine what content to create to maximize their revenue. Additionally, it may benefit free speech advocates who may believe content is being unfairly or unlawfully censored.

### *1.3. Paper Organization*

The rest of the paper is organized as follows: Section 2. provides the background of the YouTube moderation and censorship and brief monetization and its importance. The section also lists the closest research effort to that of the paper. Section 3. details the proposed methodology, i.e., data collection process, meta-data and feature pre-processing, machine learners and the process of model building. Section 4. presents the results and discussion on the performance of the machine learners. Section 6. report the limitation of the proposed work. Section 7. and concludes the paper and lists several research avenues as part of the future work.

## 2. Background and Related Work

### *2.1. YouTube Moderation and Censorship*

YouTube moderates content published on its platform through its "Community Guidelines and Policies"[1]. The Community Guidelines specify certain types of content that are prohibited from publication on YouTube's platform. The types of content fall into five categories: (i.) Spam and Deceptive Practices; (ii.) Sensitive Content, (iii.), Violent or Dangerous Content; (iv.) Regulated Goods (e.g., firearms) and (v) Copyrighted Content.

The segment of prohibited content with the most controversy surrounding it is YouTube's "Violent and Dangerous Content" category. YouTube has defined this category to include the following types of content: (i.) Harassment and cyberbullying; (ii.) Harmful or dangerous content; (iii.) Hate speech; (iv.) Violent or graphic content; (v.) COVID-19 misinformation. Most of the controversy surrounds YouTube's definition or categorization of "harmful or dangerous content." YouTube's "harmful or dangerous content" label has been used to remove or censor large swaths of content from its platform. The label has also been used to purge and censor content ranging from content related to cannabis [2] to cryptocurrency[3]. However, YouTube's purging or censoring of content deemed "far-right" or "ideologically right" has had the most outcry [4].

Indeed, critics of YouTube have alleged that it uses its "harmful or dangerous" label to remove and censor

ideological right content, which YouTube and its supporters have claimed to be "far-right conspiracy theories." Concerning the way YouTube censors content for what it deems violations of the Community Guidelines. YouTube's primary method of moderation and censorship is through demonetization [5]. YouTube's demonetization of content can come in two forms: (i) demonetization of an entire channel; or (ii) demonetizing on a video-by-video basis.

### *2.2. YouTube Monetization and its Importance*

YouTube's monetization of content is done almost exclusively through advertisements played on content uploaded by channels and content creators. Specifically, when a viewer plays a monetized video on YouTube, advertisements will appear in some combination: (i) before the video plays; (ii) during the video (called "mid-roll" advertisements; or (iii) in banners next to the video.

Fig. 2 is an example of an advertisement for "X Box Series X" playing before a video uploaded by popular YouTube content creator Mr. Beast. Portions of the advertising revenue YouTube earns by playing to video on uploaded content are paid to the content creator through the YouTube Partnership Program. Thus, the generation of advertising on published content and videos has become extremely important to content creators and has accelerated content creation on the YouTube platform. Concerning how YouTube deems content worthy of monetization, the platform has strict guidelines for a channel to become "monetized." Specifically, a channel must meet the following guidelines before content published on the channel can be monetized (i.e., have advertisements played on its content). These guidelines are: (a) 1,000 subscribers, (b) 4,000 public watch hours on the channel's content over 12 months and (c) The channel must follow the YouTube Community Guidelines. Additionally, YouTube will also evaluate other factors in monetizing a channel. While less than clear, the platform has given stated that it considers the following when making a monetization determination for a channel: (a) The channel's "Main Theme"; (b) The channel's most viewed videos and (c) The channel's newest videos and (d) The most significant proportion of watch time on a channel's videos and Video metadata (including titles, thumbnails, and descriptions). Thus, while a channel may meet the qualifications of 1,000 subscribers and 4,000 public watch hours to be monetization, YouTube holds ultimate discretion in monetizing a channel by effectively reviewing the content on the channel. After a channel is monetized, the guidelines for whether an individual video is monetized become even more nebulous. Indeed, YouTube has not published clear guidelines on what it looks at in determining whether to monetize (or run advertisements) on an individual video.

However, it is important to note that the decision as to where to monetize a video appears to be automated by YouTube. An uploaded video is typically given an instant decision as to whether it is monetized or demonetized. However, if the automated algorithm demonetizes a video, a content creator can request a manual review.

Monetization of a video is extremely important. As stated above, the earning potential drives content creators to publish content. More importantly, though, demonetization of content is an important tool YouTube uses for moderation and censorship of content published on its platform.23 Indeed, the demonization of videos has been called a form of censorship as it has several negative impacts both for content creators and users:

- The lack of ad revenue disincentives content creators from publishing certain types of content.
- Demonetization of a video affects the video's search ranking or search result on YouTube. Demonetization content is less likely to appear in searches.
- The demonization of a video affects whether YouTube will recommend that video to users. Thus, it becomes harder for users to find the content and more challenging for channels with demonetized content to grow a user base.
- A demonetized video negatively impacts a channel overall as it is generally considered a "black mark." It may affect whether a channel's other videos are monetized and therefore searchable and recommended.

Put simply, the demonetization of videos and content on YouTube suppresses that content and drastically affects a user's ability to find the content and negatively impacts a content creator in trying to get their content to an audience. Monetization and demonetization of videos play a significantly important role in YouTube on what content a user has the ability to find and consume on the platform. In essence, demonetization can act as a proxy for

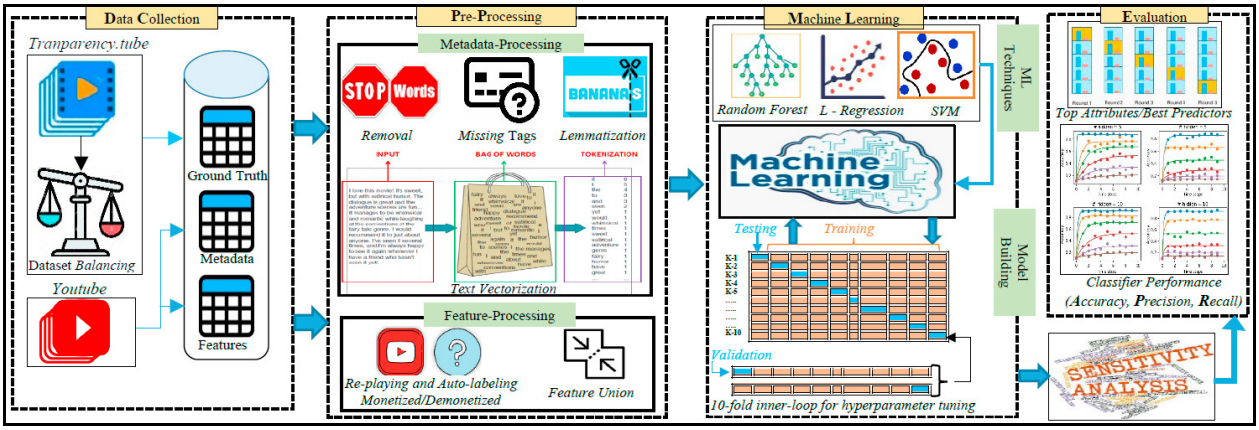


Fig. 3. A high-level overview of the proposed methodology

Table 1 . Features Derived from YouTube

| No. | Features | Description |
|---|---|---|
| 1. | Channel Description | The description of the channel the video was published on, which was supplied by the content creator. |
| 2. | Channel Subscriber Count | The number of subscribers of the channel the video was published on. |
| 3. | Channel View Count | The total number of views for all videos on the channel the video was published on. |
| 4. | Channel Video Count | The number of videos on the channel the video was published on. |
| 5. | Title | The title of the video supplied by the content creator. |
| 6. | Description | The description of the video supplied by the content creator. |
| 7. | Like Count | The number of likes on the video. |
| 8. | Dislikes Count | The Number of dislikes on the video. |
| 9. | Like Ratio | A derived value by taking likesCount and subtracting dislikesCount. |
| 10. | Comment Count | The number of comments on the video. |
| 11. | Views | The number of views on the video. |
| 12. | Duration | The length of the video. |
| 13. | Tags | The metadata tags associated with the video supplied by the content creator. |

censorship and suppression of content on YouTube's platform.

### *2.3. Related Work*

A few works have used machine learning to understand content censorship factors, especially on social media. Due to space constraints, we list the research efforts that coincide the most with our work.

*Jin Li* (2015) [6] used machine learning techniques to try to determine whether certain published "microblogs" would be deleted or censored. Li focused on the Chinese microblogging platform Weibo. He employed Naïve Bayes to classify content microblogging content and achieved an approximately 95% accuracy rate. In contrast, our work employs both textual and numeric data. As well, use classifiers other than Naïve Base.

Likewise, a recent research effort from *Cross et al.* (2019) [7] was aimed to develop a model to replicate censorship filters on Weibo (a Chinese microblogging platform) and achieved a 96.2% accuracy rate. They also developed a model to transform posts to bypass censorship filters and achieved a 30% success rate. In creating the classification models, they used MNB, SVM and Logistic Regression models. Cross et al.'s work will serve as a comparative ground for measuring the performance of our work.

*Tanash et al.* (2015) [8] attempted to detect censorship trends on Twitter through users "following" each other, focusing on "influential" users. Unlike our work, Tanash et al. focus on the followers/following network rather than the actual content being censored.

## 3. Methodology

Fig. 3 shows an overview of the proposed methodology and consists of the following steps:

### *3.1. Data Collection*

In order to choose a proper representative dataset where the bias for political ideology could be validated by our methodology, pre-classified videos from left-wing, moderate and right-wing were needed. Transparency.Tube [9] is a website that has employed machine learning techniques to classify YouTube channels according to their content. Specifically, as shown in Fig. 1, Transparency.Tube classifies YouTube channels into a handful of categories: Mainstream News, Anti-Woke, Social Justice, Left, Far Left, Right, and Far-Right. We chose the Transparcey.Tube data source because its classification of videos is generally accepted in the community and appears reasonably accurate. A diverse but balanced dataset containing 400 seed-videos were created using a protocol consisting of:

(a). Randomly selected videos from ten channels from each category on Transparency.Tube.
(b). The same number of "Left" and "Right" channels were selected in order to balance the dataset
(c). Five most recent videos from each channel.
(d). Ensured a broad coverage of video types, i.e., videos that varied in duration, long-form, recorded podcasts, short clips, and live streams.

Next, a set of features listed in Table 1 and user comments from the 400 seed-videos of YouTube were using YouTube's proprietary API.

### *3.2. Pre-processing*

**Metadata-pre-processing:** The machine learning algorithms used in our methodology, i.e., Random Forest, Liner Regression and SVM cannot be applied directly to the corpus of user comments collected for the videos. This is because most user comments are typed from smartphones and contain typos, slang words, jargon, and abbreviations. Therefore, the following pre-treatment steps are performed. (**1**) *Noun, verb and adjective extraction* – the part of speech (POS) tagging functionality of the Natural Language Toolkit, NLTK4, is used for identifying and extracting the nouns, verbs, and adjectives in the student reviews. (**2**) S*top-word removal* – stop-words in student comments are removed to eliminate terms that are very common in the English language (e.g., "and", "this", and "is"'). The standard list of stop-words provided by 'Lucene5' is used. The list is also expanded to include common words in these user reviews but are not used to describe features. (**3**) *Lemmatization* – The Natural Language Toolkit (NLTK) will not count different forms of a word (test, testing, tests, tested) together naturally, so these were simplified using the NLTK's WordNet Lemmatizer to return each word in its root form. (**4**) *Vectorization* – when querying the YouTube API for the video "tags" data, the video tags were returned in an array. Each tag was placed in an index in the array. Thus, a small script was run to turn the array into a single string with spaces separating each tag. This was important to be able to more easily use the "tags" data in the models with the "Bags of Words" approach. Converting the tags array to a string also did not impact or skew the model since using the "Bag of Words" approach meant that the individuals tags were tokenized and vectorized anyway.

**Feature pre-processing:** (1) *Auto-Labelling* – A custom script was writing using Selenium (a package in Python). The script looked for embedded HTML tags when playing a video in the dataset to determine if advertisements were present and/or played, then the video was deemed to be monetized by YouTube. However, if no advertisement were present or played in the video, then the video was deemed demonetized and labeled accordingly.

### *3.3. Machine Learning*

Our work aims to understand if a set of YouTube features (i.e., attributes) can predict whether a change in the values of the attribute will lead to (demonetization) censorship of the video.

Table 2. Significant Features and Their Ranking

| No. | Features | Feature Ranking | | |
|---|---|---|---|---|
| | | Logestic Regression (LR) | Support Vector Machine (SVM) | Random Forest (RF) |
| 1. | Channel View Count | 0.035955 | 0.027778 | 0.047191 |
| 2. | Channel Subscriber Count | 0.033708 | 0.041667 | 0.006742 |
| 3. | View Count | 0.000000 | 0.000000 | 0.000000 |
| 4. | Like Ratio | 0.000000 | 0.000000 | 0.000000 |
| 5. | Dislike Count | 0.000000 | 0.000000 | 0.000000 |
| 6. | Like Count | 0.000000 | 0.000000 | 0.000000 |
| 7. | Comment Count | 0.000000 | 0.000000 | -0.00224 |
| 8. | Right | 0.000000 | 0.000000 | 0.000000 |
| 9. | Left | 0.000000 | 0.000000 | 0.000000 |
| 10. | Center | 0.000000 | 0.000000 | 0.000000 |
| 11. | Duration | 0.000000 | 0.000000 | 0.000000 |
| 12. | Channel Description | 0.000000 | 0.000000 | 0.026966 |
| 13. | Tags | 0.000000 | 0.000000 | 0.020225 |
| 14. | Title | 0.000000 | 0.000000 | 0.000000 |
| 15. | Description | 0.000000 | 0.000000 | 0.000000 |
| Legend | 1st Rank | 2nd Rank | 3rd Rank | 4th Rank |

More simply, to find if the set of attributes can act as a proxy to the YouTube Blackbox censorship algorithm that heavily relies on demonetization to censor a video. We model our work as a classification problem where change in the features values can fall into one of the two classes: Monetized (YES) or De-Monetized (NO).

***Machine Learners:*** Several machine learning techniques exist in the literature, such as Neural Networks, Naïve Bayes, and Stochastic Gradient Descent suitable for solving the classification problem. In our proposed methodology as shown in Fig 3., we selected four learners: C 4.5 - a simple decision tree classifier, Random Forest - an advanced decision tree classifier, Linear Regression (LR) and Support Vector Machine (SVM). The rationale of choosing the learner for our methodology is based on the fact that: (a) Decision Tress produces explainable models. The explainable models facilitate understanding the video monetization and demonetization phenomenon. As well, to find the essential features that are likely to be the proxy candidate for YouTube; (b) Linear Regression (LR) is easy to interpret and has a considerably lower time complexity when compared to some of the other machine learning algorithms; (c) The C 4.5, RF and SVM facilitate performing sensitivity analysis to determine the most important features that are likely to be the proxy to the censorship algorithm [10][11].

***Model Building:*** All the machine learners take features listed in Table 1 as input. Most of the features were selected based on YouTube's statements of what factors the platform considered in monetizing. A few more features (not listed in Table 1 which the author believed may play a role or influence the YouTube monetization algorithm, i.e., (duration (favouring longer videos over shorter videos), tags, left, right, centre, like count, dislike count, and like ratio). The target variable i.e., class variable was "monetization," which was a 1 (Yes), if a video is monetized and a 0 (No), if the video is demonetized.

Because the feature set contained both numeric and string text, the creation of the models was complicated. More precisely, the string or text features needed to be reduced to numeric form. To accomplish this, the "Bag of Words" method was employed. The "Bag of Words" approach tokenizes and vectorizes text data. For each feature, an array is created. Words or tokenized phrases are then turned into or assigned numbers in the arracy.This method then allows the frequency with which a word or tokenized phrase appears to be quantified and placed in the model. Additionally, "stop words" such as "a," "the," "of," etc. are removed. After the string-based features were vectorized, a "Pipeline" was created in order to merge the numeric and string-based features in the models. In scikit-learn, this is achieved by a function called a "Feature Union."

In each model, the dataset was divided into a training set and a testing set. The models were trained on the divided training set. Once a model was generated, the model ran on the testing set and checked for accuracy. After running the four models, i.e., C 4.5, RF, LR and SVN, it was noticed that the vast majority of the features were not significant in any of the models and are listed in Table 2. Specifically, the features "like count," "dislike count," "like ratio," "comment count" and "view count" were not significant in any of the models and make sense. The decision to monetize a video is made when a video is uploaded using an automated algorithm by YouTube. Data related to views, likes and comments is generated after the video is published and a decision to monetize or demonetize a video has already been made by YouTube. Thus, the fact that these features showed no significance across all three models is consistent with YouTube's process in making decisions to monetize or demonetize videos.

## 4. Discussion of Results

The ***Logistic Regression (LR)*** model achieved an accuracy of predicting whether a video would be monetized of approximately 70%. The two most significant features in determining whether a video is monetized was *duration* and *channel subscriber count*. The recommendation of the two important features is appropriate, i.e., The longer the video, the more advertisements YouTube can play on it, the more likely the video would be monetized. Likewise, the more subscribers a channel had, the more likely the video would be monetized. More importantly, the relationship between *channel subscriber count* and *monetization* makes even more sense. Channels with large subscriber counts have likely benefited from YouTube recommending its videos. Videos are generally recommended when they are monetized. Thus, channels with large subscriber counts have effectively fed off YouTube's monetization algorithm and have earned goodwill with the monetization algorithm. Consequently, the relationship appears to be rational.

The ***Support Vector Machine (SVM)*** model likewise achieved an accuracy of predicting whether a video would be monetized of approximately 70%. The only two significant factors were *channel subscriber count* and *channel view count*, two highly correlated variables.

The ***Random Forest (RF)*** model achieved the highest prediction accuracy, i.e., 85% whether a video would be monetized. As expected, the basic tree learner, i.e., C 4.5 achieved the least prediction accuracy, i.e., 67 %. Due to its low accuracy and space constraints the significant features proposed by C 4.5 are not listed in the Table 2.

The four prediction models generated suggest that the political ideology of a YouTube channel– whether it be left, right or center – has no significance in YouTube's algorithm for determining monetization of an individual video. This is indicated by the fact that in all the four models, the features related to political ideology (e.g., "left", "right" and "center") had no significance. This finding appears to undermine repeated accusations that YouTube "censors" right-leaning content. If YouTube favors one political ideology over another, it does not appear to be playing a role in making the determination to monetized or demonetize content on the platform.

## 5. Machine Learner Performance Evaluation

### *5.1. Logistic Regression.*

We used the Receiver Operator Characteristic (ROC) cure as an evaluation metric, well suited for binary classification problems. The average performance value of 10-folds of cross validation is 0.7012. It took 15 seconds to train the model. Significant improvement was noticed, i.e., 0.8109 after the application of 10-fold cross validation. The time overhead incurred for the cross validation was 3 minutes.

### *5.2. C 4.5 Decision Trees.*

The decision tree classifier was built using 10 nodes and the depth of the tree was limited to 5. The number of nodes used in the model is 15 and the depth of the tree is 3. The performance metrics used is Gini index. The average performance value of 10-folds of cross validation is 0.5344. It took 175 seconds to train the model. Not noteworthy improvement was noticed after the application of 10-fold cross validation. The time overhead incurred for the cross validation was 15 minutes.

### 5.3. Random Forest

The RF classifier was build using 10 nodes and the depth of the tree was limited to 5. The performance metrics used is Gini index. The average performance value of 10-folds of cross validation is 0.7411. It took 15 seconds to train the model. Remarkable improvement was noticed after the application of 10-fold cross validation, i.e., 0.9111. The time overhead incurred for the cross validation was 16 minutes. So we see that the model, based on the random forest algorithm, shows the best performance so far. However, it takes the most time to train the model, as the algorithm is more complicated.

### 5.4. Support Vector Machine

The performance metrics used for SVM are based on a confusion matrix. In case of SVM, it took 240 seconds to train the model. It is important to note, that we are relying on limited videos from Transparency. Tube (as ground truth), hence we are dealing with unbalanced classes Monetized and demonetized). As a result, 65% videos that were actually monetized were correctly classified with an accuracy of 92%. Whereas 45% of videos that were De-monetized were classified with an accuracy of 76%.

## 6. Limitations of the Work

The major limitations of the proposed work are listed below:

1) The research was conducted in the USA on videos & channels meant for primarily an American audience. A Mexican or Canadian viewer many not even see an advertisement on the video.
2) Frequency of advertisements —the paper deems a video to be "monetized" in the presence of advertising. The paper does not distinguish the frequency of advertisement in a video.
3) YouTube's monetization algorithm is built with advertisers in mind which can act as a form of censorship-by-proxy against creators. Thus, creators self-censor their content, attempting to work within the algorithm by mature themes from their content, or work against the algorithm by deliberately posting content they know will be censored, and posting a companion video which will not be censored to inform viewers of the censored content and provide a direct link to it. The work does not distinguish such videos nor were they part of the collected data set.

## 7. Conclusion and Future work

The paper presents a machine learning methodology that facilitates understanding if a set of YouTube features (i.e., attributes) can predict whether a change in the attributes' values will lead to (demonetization) censorship of the video. More simply, to find if the set of features can act as a proxy to the YouTube Blackbox censorship algorithm that heavily relies on demonetization to censor a video. Using a set of four machine learners, i.e., C 4.5, Random Forest, Linear Regression and Support Vector Machine, the methodology achieved up to 87 % of the prediction accuracy.

As future work (**FW**) future, we want to expand the methodology along several avenues, included but not limited to:

***FW-1****:* We plan to consider

(a). if the number of advertisements in a video is a proxy candidate for censorship and

(b). frequency of advertisement in different markets, i.e., Canada and Mexico, changes since YouTube may have different advertising standards for different markets.

***FW-2****:* We also plan to expand our research across various demographics to:

(a). categorize user experience across multiple categories (i.e., News, Sports, Drama, and Finance) and

(b). identify how important factors/attributes, across categories, differ for monetization and demonetization.

***FW-3:*** Along with our industrial partner, we also plan to give more credibility to our findings by working alongside small content creators. So that additional data is available, and that the monetization status of videos can

be conferred, rather than solely relying on our custom video replaying and labeling script.

***FW-4:*** Our methodology uses NLP techniques (as part of pre-processing) on user comments to identify topics (i.e., nouns) discussed in the user comments. However, in our pilot study, we have not yet included these topics in the model construction. We plan to expand our work by

(a). incorporating aspect-based sentiment analysis on dominant topics of every video and

(b). find the magnitude of the polarity of the sentiment, i.e., user option of each topic. Enumerating the sentiment and polarity of user comments into the model(s) is likely to increase its performance.

## References

[1] 9-YouTube, "Rules and Policies, Community Guideliens", aviable at: YouTube Community Guidelines & Policies - How YouTube Works, downlaoded 2nd January, 2021.

[2] "YouTube's Cannabis Content Purge 'A Huge Loss of Cultural History," CannCentral, available at: https://www.canncentral.com/youtubes-cannabis-content-purge-a-huge-loss-of-culturalhistory. Downloaded at 2nd January, 2021.

[3] Frobes, "YouTube Goes to War with BitCoin and Crypto," available at https://www.forbes.com/sites/billybambrough/2019/12/26/googles-youtube-goes-to-war-with-bitcoinand-crypto-update/?sh=5d1368f71b63.Downloaded 3rd Janudary, 2021.

[4] The New York Times, "YouTube Cracks Down on Fair-Right Videos as Conspiracy Theories Spread," available at: https://www.nytimes.com/2018/03/03/technology/youtube-right-wingchannels.html. Dowloaded 3rd January, 2021.

[5] Francesca Duchi , "Problematic Algorithms: YouTube's Censorship and Demonetization Problem,", available at https://medium.com/future-vision/problematic-algorithms-youtubescensorship-and-demonetization-problem-f1634a63d1b4. Downloaded, Jnauary 2nd, 20201.

[6] Jin Li, " Predicting Large-Scale Internet Censorship – A Machine Learning Approach", Thesis, aviable at: https://libra2.lib.virginia.edu/downloads/d504rk52c?filename=Jin_Li_MS_Thesis.pdf. Downloaded 9th January 2021.

[7] Christopher Cross, "Bypassing Censorship Reverse Engineering a Social Media Censorship Classifier to Generate Adversarial Posts", in procedings of Cross BypassingCR, 2019.

[8] Rima S. Tanash, Abdullah Aydogan, " Detecting Influential Users and Communities in Censored Tweets Using Data-Flow Graphs", 2016.

[9] Webstite: https://transparency.tube/. Last accesed January 21st, 2021.

[10] David Baehrens, Timon Schroeter, Stefan Harmeling, Motoaki Kawanabe, Katja Hansen, and Klaus-Robert Mü ller. 2010. How to Explain Individual Classification Decisions. JMLR , Vol. 11, 2012.

[11] Bouaziz A., Dartigues-Pallez C., da Costa Pereira C., Precioso F., Lloret P. (2014) Short Text Classification Using Semantic Random Forest. In: Bellatreche L., Mohania M.K. (eds) Data Warehousing and Knowledge Discovery. DaWaK 2014. Lecture Notes in Computer Science, vol 8646. Springer, Cham. https://doi.org/10.1007/978-3-319-10160-6_26