

Exclassifier: A Novel Technique for Detecting Extremist Videos in Social Media

Anuradha Pillai¹ and Prachi Kaushik² ¹Assistant Professor, CE Department, YMCAUST Faridabad, India Email: anuangra@yahoo.com ²PG Student, CE Department, YMCAUST Faridabad, India Email: prachikaushik.4@gmail.com

Abstract—With the growing popularity of social media, social network (you tube) remains the largest as well as the most popular video sharing site. However, terrorists groups have made YouTube as a focal point for targeting innocent and vulnerable people. They propagate their ideologies to mainstream audience who otherwise would not visit their website. Hence, there is a need to detect such videos to prevent online radicalization among the users. The paper proposes a metadata and audio based classification method for detecting such videos which promote hate and violence by mining the user generated metadata such as title, description which the uploader of the video adds along with finding patterns to classify an audio into violence class such as gunshots and screams.

Index Terms-Social Media privacy, web security, extremism detection.

I. INTRODUCTION

As social media is growing very rapidly, people are spending significant amount of time on these websites. Specifically some social networking sites such as Facebook, Twitter, YouTube, and LinkedIn have become the most used way of interaction among the people in the worldwide.

YouTube is one of the most popular social networking website where users easily upload the video, watch the video and share videos on other social networking websites. YouTube have no limitation on number of videos users can watch, upload and share.

According to the YouTube statistics:

- 1. Over 1 billion unique users visit YouTube each month[9];
- 2. About 6 billion hours of video are watched each month on YouTube [9].
- 3. 100 hours of video are uploaded to YouTube every minute [9].

This statistics [9] shows the high popularity of YouTube on internet also, users can comment in textual form, subscribe for any channel, can search videos on keyword & category, like or dislike the videos. Since YouTube allow easy upload and downloading of videos, thus some users use YouTube to spread Hate and extremism among people by uploading videos having hurtful and provoking contents. These videos are not allowed according to YouTube privacy policy. Further these videos waste the bandwidth for the users who are not willing to watch these videos.

Since, there is no proper mechanism for automatic identification to detect objectionable videos that a user is uploading; extremist groups put forth hateful speech, offensive comments and messages focusing their

Grenze ID: 01.GIJCTE.2.1.23 © Grenze Scientific Society, 2016 mission. This material is used to facilitate recruitment, gradually reaching worldwide viewers, connecting to other hate promoting groups, spreading extremist content and forming their communities sharing a common agenda. The presence of hate producing users in large amount is major concern of YouTube, government and law enforcement agencies [2]. Even after many community guidelines and administrative efforts made by YouTube, there are still huge amount of extremism and crime producing videos on YouTube. Because identifying these Hate producing videos became technically challenging problem [2], a solution for this challenge is needed to combat and counter online radicalization. We categorized the data into two parts: training data and testing data. In order to identify, we extract the metadata of YouTube videos such as video-id, title, likes, dislikes, description etc. Training data is used for calculating threshold value (with the help of manually created lexicon lists) while testing data is compared with calculated threshold. The audio based classification is done as some users may write misleading titles and description due to which the videos will be difficult to detect. Hence, combination of the meta-based and the audio based classifier is an effective solution to such a problem.

II. RELATED WORK & RESEARCH CONTRIBUTIONS

This section gives an overview of the previous works in the area of detecting extremists and violent content online.

- H. Chen et al discuss the classification of online videos by extraction of user-generated text data title. Meta-description & comments and three types of text feature (lexical, syntactic, contentspecific features). Extremist videos were classified using three classification techniques C4.5, Naïve Bayes and SVM. According to experimental results SVM performed better than C4.5 and Naïve Bayes classifiers [1]. But this technique does not involve the non-text features of video such as audio, video, colour, texture and motion vectors.
- 2. A. Surekha et. al proposed an approach to discover hate and extremist videos, central and influential users and communities from YouTube by data mining as well as social network analysis techniques using friends, subscription, favourites and related video concept [2].
- 3. W. Chung in their case study of jihad on web developed a useful methodology to collect and analyse dark web information. Terrorist clusters and their use of web are identified using Information Visualization such as multi-dimensional scaling and snowflake visualization. This methodology was applied on 39 jihadist websites and developed visualization of site contents, relationships and activity levels [3]. This method is not scalable and it doesn't take into account the volatile nature of jihad websites.
- 4. A. Bermingham exploits the potential for the online radicalization by using social network analysis. Sentiment analysis of profiles and comments suggests that female users have most extreme and less tolerant views. Lexical analysis suggests that two most frequently used words are Allah and Islam. Social network analysis indicates an increased leadership role of women online. According to study the topics of discussion on the YouTube groups were mainly on America, Christianity, Islam, Israel, al-Qaeda [7]. The breath of the corpus is small and the lexicon used for sentiment analysis is not domain specific for the problem. It does not take into account the non-english text for sentiment analysis algorithm
- 5. T. Chalothorn used sentiwordnet to detect the opinions and emotions for the radical web forums by assigning positive and negative score to each wordnet. Words were stored on a BOW (bag of words and POS (part of speech) was used for tagging words. By sentiment analysis of sentence score it was concluded that Qawem web forum has more radical content than montada web forum [8]. This approach is a comparative study of only two web forums, other radical web forums are not considered.
- 6. T. Giannakopoulos et. al proposed a methodology to detect violent scenes in movies using twelve audio features and visual features combined together. The video features included certain motion specific features such as average motion, motion oriented variance and detection features for the face detection in the scenes. The performance of the system is 83% and only 17% of the scenes are not detected.[11]
- 7. Xingyu Zou et. al in this paper proposes a text, audio, visual based violence content classification. The first stage is a text based classifier to identify potential movie segments. The second stage used a combination of audio and visual cues to detect violence. Audio features like audio energy, energy

entropy and visual features like motion intensity, colour of the flame, bleeding and shot lengths are extracted to enhance the classification task.[12]

- 8. L. Gerosa et. al in their approach trained two parallel GMM classifiers to differentiate gunshots and screams from noisy environment which belong to the class of violence. A set of 47 audio features were used for the classification and the proposed system guarantees a precision of 90 %.[13]
- 9. T. Giannakopoulos, D. Kosmopoulos used frame level time domain and frequency domain features along with the SVM classifier to detect violence content. The recall rate was 90.5% which could be further improved by MFCC coefficients.[10]
- 10. E. Vozarikova et. al presents a methodology to detect dual gunshots in noisy environment using features such as MFCC, MELSPEC, skewness, kurtosis and ZCR. The combination of different features were evaluated by the HMM classification technique.[15]
- 11. Pikrakis identified gunshots by dynamic programming and Bayesian network. The posterior probabilities were calculated by combing the decisions from a set of Bayesian network combiners and 80% of the gunshots were correctly detected.[14]

III. PROPOSED WORK

Proposed work in this research proposes a novel technique (Exclassifier) to detect hate and violence producing videos. Architecture of Exclassifier is shown in figure1. The whole process is divided into 4 main components: Metadata based classification, Audio based classification, Analyzer and Final classification. Working of various components of the proposed system is discussed in next sections.



Figure 1. ExClassfier (Extremist video classifier)

A. Metadata based classification

The first component of the system interacts with social media site (i.e YouTube in our case). The detailed discussion of all components is given in next session.

B. Video Collection

For the training purpose, the task of data collection is done manually by querying the YouTube search interface using the keywords hate, behead, kill, bomb etc. The video ids for the videos which produce hate and extremism are collected by inspecting the query result.

C. Metadata extraction

This module extracts user generated text data for each video id obtained from the video collection step, using YouTube API. The linguistic based features such as the title and the description added by the uploader of the video along with duration of video, category, ratings, number of comments, likes and dislikes are extracted.

D. Prepare training dataset

The training dataset is prepared using video collection and metadata extraction step discussed above. The extracted metadata for each video id is added to this set .The outline of the training dataset is shown in Fig 2.

	Vid	Title	Description	category	duration	comments	likes	dislikes	rating	
--	-----	-------	-------------	----------	----------	----------	-------	----------	--------	--

Figure 2: Training dataset Data-Structure

E. Data Pre-processing

The extracted title and description undergo a preprocessing step which consists of two steps:

1. Tokenization:

The process of tokenization involves breaking a stream of text into words, phrases or meaningful elements called tokens. The text of title and description are broken into tokens {t1, t2... tn} taking the whitespace character as a delimiter.

2. Stop word removal:

Stop words are the most common words which have little value in classification process. This step filters stopwords using a list of standard English stop word list. The stop word list comprises of articles (a, an, the), prepositions (to, from, on, by etc), pronouns (I, we, your, us etc)

F. Construction of Hate Term Lexicon & People Term Lexicon

The two lexicons (Hate term list, People term list) are being constructed in this step. The construction of initial hate term lexicon is done by adding the most frequent hate terms involved in the training dataset. The lexicon list grows in size by adding synonyms as well as hyponyms for the existing terms.

The people type lexicon involve the proper nouns i.e. the name of the terrorist groups, jihadist, words like Muslim, Hindu, etc and the people who spread jihadist ideologies. Initial hate term and people term lexicon lists are shown in Table 1.

TABLE I. LEXICON LISTS

Hate term List	People term List
Attack, behead, bomb, threat, terror, kill, jihad etc	Osama, Muslim, Hindu, Isis , sayeed , lashkar, Allah



Figure 3. Score Calculation

F. Matching Function

The tokens for the title (T_1) and the description (D_1) are compared with the lexicon L1 (Hate term lexicon) and L2 (People term lexicon).

The inputs to the matching function M (T_t, D_t, L1, L2):

- T_t: Title tokens
- Dt: Description tokens
- L1: Hate term lexicon
- L2: People term lexicon

The matching function returns a similarity score (S_i) for each video- id *i* in training set. The similarity score shown in Figure3 is calculated as follows:

For the video-id *i*,

TL1 = number of title tokens (T_t) matching lexicon L1

TL2 = number of title tokens (T_t) matching lexicon L2

- **DL1**= number of description tokens (D_t) matching lexicon L1
- DL2 = number of description tokens (D_t) matching lexicon L2
 - 1. Similarity score S(Title)

- S1 = $TL1/_{Tt}$ (Ratio of hate terms in title) S2 = $TL2/_{Tt}$ (Ratio of people terms in title) 2. Similarity score S(Description)

 $S3 = \frac{DL1}{Tt}$ (Ratio of hate terms in description)

 $S4 = \frac{DL^2}{Tt}$ (Ratio of people term in description)

G. Calculate Threshold

The threshold of training set is calculated using the score S assigned by the matching function M.

The arithmetic mean of the scores is computed for title and description shown in table 2.

The threshold value for each type is calculated. The values obtained by the experimental results are shown below in Table 3 where, Type 1: hate terms Type2: people terms.

Title	Description
$th1 = \sum S1_i/N$	$th3 = \sum S3_i/N$
$th2 = \sum S2_i/N$	$th4=\sum S4_i/N$

TABLE III.	THRESHOLD	VALUES
------------	-----------	--------

Feature	Type 1	Type 2
Title	.21(th1)	.13(th2)
Description	.10(th3)	.05(th4)

H. Classification

The classification of the videos into extremist class or unknown class is done by training the one-class classifier with the positive class examples where the positive examples are the videos that depict extremism. After training task, the test videos are checked one by one for extremism by undergoing the same process of metadata extraction, pre-processing, and matching function to calculate scores.

The scores of the test video s1, s2, s3, s4 (section 3.4) are compared with th1, th2, th3, th4 (threshold values from Table II and Table III) as shown below.

- 1. If $(s1>th1 \parallel s2>th2)$ and $(s3>th3 \parallel s4>th4)$ we classify video as *extremist video*.
- 2. If $(s_1 > th_1 || s_2 > th_2)$ or $(s_3 > th_3 || s_4 > th_4)$ we classify the video as *unknown class* but the content analysis of video can be used to correctly classify this video content.
- 3. If neither condition 1 nor condition 2 satisfy the video is classified as *safe video*.

IV. AUDIO CLASSIFICATION

This module of the proposed work inputs a segment of the audio and divides it into frames. The extremist videos are detected using time domain features and frequency domain features.

A. Time Domain Audio Features

Energy

Let $x_i(n)$, n = 1,...,N the audio samples of the ith frame, of length N. Then, for each frame i the energy is calculated according to (1):



$$E(i) = \frac{1}{n} \sum_{n=1}^{N} |x_i(n)|^2$$
(1)

Figure 4. Energy waveforms (E) for music, speech, gunshot, scream

The extracted energy feature is used to detect silent segments in the audio signal as well as to differentiate between audio classes. The variation of energy(CV) in the speech segment is higher than music signal as its energy alternates from high to low. The statistics calculated for energy is the CV(coefficient of variation. Energy waveform (E) of (a) music (b) speech (c) gunshot (d) scream is shown in figure 4 shown above. According to the CV values the order of audio signal is is Music < scream < speech<Gunshot. Gunshot has the highest value for CV and music the lowest CV.

B. Zero Crossing Rate

Zero crossing rate (ZCR) is the measure of the number of times the signal alternates from positive to negative and back to positive. The ZCR value of periodic signal is less as compared to noisy signal. This feature has been used extensively to identify speech and music segments. The formula (2) to calculate ZCR is given below:

$$Z(i) = \frac{1}{2N} \sum_{n=1}^{N} |sgn[x_i(n)] - sgn[x_i(n-1)]|$$
(2)

The CV value of the ZCR sequence of the speech segment is higher than the music segment due to abrupt changes from positive to negative. Statistics calculated for ZCR is CV and Mean.

The Figure 5 depicts the waveform for the ZCR values for Music, speech, gunshot, scream. According to the experimentation value of ZCR_{CV} is in the following order: - Scream< Music < Speech<Gunshot. The highest value of ZCR_{CV} is for gunshot and the lowest is for Scream. The mean value of the zero crossing rate is computed and according to this statistics if we arrange the series in increasing order of mean values, the order is:-Music <speech <scream < gunshot.

C. Energy Entropy

Energy entropy is a time domain feature which measures abrupt changes in the energy level of an audio signal. It is calculated by dividing each frame into K fixed duration sub-frames. The energy e_j^2 is calculated (3) for each sub-frame by dividing the sub-frame's energy, by the whole frame's energy.

$$e_j^2 = \frac{Esubframe_j}{Eshort\,frame_i} \tag{3}$$

The energy entropy of the sequence of frames (4) is calculated by the noramalized energy e_j^2 calculated in (3).

$$En(i) = -\sum_{i=0}^{K} e_j^2 \log_2(e_j^2)$$
(4)

The statistics value of the energy entropy is taken as the coefficient of variation. According to the experimentation the audio signals with abrupt changes has a higher value for CV. Gunshots and speech have larger value for the coefficient of variation compared to screams and music.



Figure 5. Waveform for the ZCR values for Music, speech, gunshot, screams

D. Frequency Domain Audio Features

This domain refers to the analysis of the audio signal based on the frequency values rather than the time domain. It gives information regarding the signal's energy distribution over a range of frequencies. A time domain signal can be converted into a frequency domain by applying transforms on the signal values (FFT). The Fourier transform is a mathematical operation which converts a time domain signal into its corresponding frequency domain.

E. Spectral Centroid

It is measure used in digital signal processing to identify a spectrum. It signifies the concentration of the centre of mass of the spectrum. This feature gives higher values for the intensity of sound spectrum. Spectral centroid for screams has a low deviation and speech signals have highly variated spectral centroid. Spectral Centroid for speech, scream and gunshot is shown in the above figure 6. Gunshot has the highest CV value and Scream has the lowest CV value, hence the order is: Scream<speech<gunshot The equation (5) is given below:



Figure 6. Spectral Centroid for speech, scream and gunshot

F. Analyzer

The new audio which is to be assigned a label from the following labels {music, speech, scream, gunshots}. Before the function of the analyzer the statistics (refer Table 4) used for each feature is calculated For an audio with a sampling rate 44.1 KHz the calculated statistics value of each feature is passed to the analyzer for the classification task.

1. IF $E_{CV}(a) > 100$ && ($ZCR_{CV}(a) > 100 \parallel ZCR$ mean > 0.1000) && $En_{CV}(a) > 200$ && $C_{CV}(a) > 100 \parallel (RO_{\mu}(a) > 0.50 \parallel RO_{med}(a) > 0.50) \rightarrow$ GUNSHOT

2. IF ($E_{CV}(a) > 100 \&\& (ZCR_{CV}(a) < 100 || C_{CV} < 100)$), entropy of the audio is calculated

If entropy_{CV}> 200 && ZCR _{Mean} > 0.1000 \rightarrow GUNSHOT audio with multiple shots

3. IF $E_{CV} > 100$ && $ZCR_{CV} < 100$ audio may belong to any of the three classes {music, speech, scream} then centroid C is checked.

- 1. IF $ZCR_{CV} < 20$ && $ZCR_{Mean} > 0.060$ && C_{CV} is <10 (ZCR value is low and mean value is high centroid CV value is as low as less than 10) \rightarrow SCREAM
- If this condition does not hold go to step 4
- 4. Now two labels are left {music and speech}
 - 1. $E_{CV}(\text{speech}) > E_{CV}(\text{music})$. If Ecv < 100 audio may be a music signal
 - 2. ZCR_{CV} , ZCR_{mean} , C_{CV} is lower for music signal than speech signal
 - 3. Compare the calculated value for the audio with the vector for speech signal and music signal. The vector is represented as shown below

< ZCR_{CV}, ZCR_{Mean} , C_{CV}>

SPEECH SIGNAL: < 76.30, 0.0429, 75.72>

MUSIC SIGNAL: <57.89, 0.0299, 23.54>

Calculate the difference of the values from the respective vectors.

Percentage of silence intervals in speech is more than music. Speech contains a series of discontinuous unvoiced and voiced segments. The graphical representation of the waveform is shown in Fig 7 and Fig 8 which depicts the silence interval in music and speech signals respectively.

$$SI = \frac{\text{Number of signal values with amplitude} < 0.01}{\text{Length of signal(L)}} \times 100$$



Table IV: Statistics

Statistics CV CV, mean(μ) CV

5. The classification of audio signal into music or speech is done by using the difference of values of audio signal from the vectors and the silence interval. If difference is less for music signal && SI <3.00 →MUSIC ELSE IF Difference is less for the speech signal && SI> $3.00 \rightarrow$ SPEECH

Otherwise the audio is classified as unknown class.

V. FINAL CLASSIFICATION STEP

- 1. If from stage 1(metadata based classification) the video is classified extremist then the video is labelled as 1 otherwise 0.
- 2. The second stage classifies audio as either music, speech, scream, or gunshot
- 3. If video labelled as 1 during stage 1 contains scream or gunshot or both then according to the classifier the video is classified as extreme video Else If the audio has speech segment the classifier returns negative result concluding that the audio is a news channel video showing extremist content.
- 4. If the video labelled as 0 in the first step contains gunshots and screams is assigned the extremist class.

VI. EXPERIMENTAL RESULTS

For detecting the extremist videos, test run is carried on 50 Videos of YouTube, the experimental results are shown in Table V. Exclassfier technique is applied to all videos. Results are shown in fig 9, fig 10, and fig 11. The meta-classifier which is the first stage of the Exclassifier has a recall rate of 72 %. For testing of the audio classifier 25 audio segments containing speech, scream, music and gunshots were taken, the classifier performed with a recall rate of 76 %. The combined approach of the meta-classifier and the audio based classifier is an effective method to detect extremism video with a recall rate of the Exclassifier 84%.

Feature	Correct	Incorrect	Recall
Title(T)	33	17	66%
Description(D)	19	31	38%
T + D	36	14	72%
T+D+ audio	42	8	84%

TABLE 5. EXPERIMENTAL RESULTS

Concentration of the data base of t	inyad	ninisi pipiter	radvättikler	des takåsarv	reSitarystelitokoreEuuStfolo45a/u7729	51766645849	WURL (Assiptphthe)	TC 2-4	sk.com			٩	28	+ 1	1 I I	C C kahet horystrivit sh	do-pachilitable-des, taklisever-Ditarget-litokan-laa	69%884%32x27219x52789%6x3049%8241-15.cc/.php?x8+p	7 C 2	* Askon	P 1	18 +	f
Date Date <th< th=""><th>155</th><th>overet to 🛛 D. C.</th><th>boro pochi</th><th>1 Bidle</th><th>n<u>l</u>_d-Exit: "stores video details of video</th><th>is 👘</th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th></th><th>X</th><th>phpMuAdmin</th><th>🗕 💰 Seces mysd wanpoeres a 🕤 Dittatore p</th><th>och i 📲 İdle 1899</th><th></th><th></th><th></th><th></th><th></th></th<>	155	overet to 🛛 D. C.	boro pochi	1 Bidle	n <u>l</u> _d-Exit: "stores video details of video	is 👘									X	phpMuAdmin	🗕 💰 Seces mysd wanpoeres a 🕤 Dittatore p	och i 📲 İdle 1899					
• • • • • • • • • • • • • • • • • • •	Stro	cture 🛿 Si	L Q Se	sarch ji	Insett 👼 Export 👼 Import	Philope	s 🕜 Operations	% Triggers							1	A & O C C	E Browse 💥 Structure 📙 SQL 🔍	, Search 🖟 Insent 🚊 Export 🚊 Import	e Piv	lloges /	Operations 31 Triggers		
If is place glows (thich): Impact manual problem (thich):		₹ sid	12		descp	cat	author	duration m	ax, rat ra	at_count	ikes di	dike vie	w_count i	lav con	ment	_ GNew	+ Options	0.11		-			
File For going Show Multitude Marcine	y 0	Delete 12Ex67	1/Dc Vid 81.1	es shows attack on	Surveillance video shows a man charging at four New York Police officers with a batcher (1876). Im Scientishan mass	News & Politics	CIN	153	5	545	101	45	113460	0	367	e-g stamator_schena e-g nysd e-g petsmarce_schena	J	te token vdeo shows, ax, attack, on, sypel, efficers	68_per 0.142957	or ppi			
and logs, in the density of a solution of	7 0	Delete 3x63X/A	Biyc Ma ata	ulens ucked 76	When I have watched the video them my eyes are broken into tear, I don't know	News & Politics	Truth of Jihad	436	5	82	85	м	19001	0	130	ie-g pachi III New III Mena	Copy Obere Jobicherbryc Copy Obere Jobicherbryc Copy Obere Jobicherbryc	muslims attacked. 76 hindu tamlys 5 temples, nayakh aktaruddin owaks hate speech, 100cr, hindus, owaks,	0.161618	0.454545			
For globan Algebra de la construction For globan	Note all adapts — see a handle being out of the vision and 5 henging is not. See an second can be then. The as a leaded second on the Note and elements and the prove its implements they meth all its hender. The leader is the formation on real by reader							er yr des_tok er yr toket	Copy Ovinte BrukWyWyadd Set Ji Copy Ovinte cGQeTOsh_0	numba terocattacks, chilog fostage people trap chistians, caliphate, islamic, state, pat, 4,	0.25	0.142957											
is p-(n) g-(b) / is p-(n) g-(b) / is p-(b)							i in the second	Copy Obiete CLg4abSPSE Set 3 Copy Obiete CLg4abSPSE Set 3 Copy Obiete clmUkpHz24	safia, ahmed jetionieu, pikih, to tenseist, hindus, killed, christians jedia	0 4.0	0												
Bit Prop Bit Prop					group burnt their houses stemples. They state everything. Nathing is left. The hindus are tatally helpless in												Office and the set of the se	new video first, place, bit tower, S. 11, S. 11, Serons	0.166667	0.0555556			
Exp Exp @ 2046 Matchino Sass and Head Tools as and Head Head Sass and Head Head Head Head Head Head Head Hea					Bangladesh. About 47 houses and 6 temples ar												📋 🥜 Edt 🕌 Copy 🤤 Delete ETqEsLVTDpo	jsumalist, danger, kabul	0 333333	0 333333			
Sector (a) / (b) / (7 0	Delete TukTilu	DiA 1520	baruddin aisi Hate	Akbaruddin Owaisi is an Indian politician and a NLA. He belongs to All India	People & Blogs	Dlip Sani	130	5	791	355	435	122292	0	339		FI 2 Edt Si Copy Collete CCeCeLCOW	syna/sa,atlack, synan,army,base torture.cu,ontariamo,cu,debook	0	-			
Name J (if is (ice) give in the Victure is strated or biseduce to bisedu			Spi Aga	ech anst 100cr	Majis-e Itihad al-Muslimin and is its foor leader in Andrea Pradesh Legislative												Est H Copy Online H2tzVpGpd	death threat, dawahlims, goes, doc, hurting, minor,	0.125				
Image by the stratement base, My detail training J ≤ B ≥ Cog @ Shim W000024er (# 2 yieldment and a plantaney detail 1 mm Emmi 100000000000000000000000000000000000	Hindus	Aus Assentry		📋 🥜 Edt 🙀 Copy 🤤 Delete Ind-OQuelos	salam shuhada hum,koi shahadat xahi,bheelay	0	0																
1 2 7 458 3 1 1 1 2 7 458 3 1 1 1 1 2 2 3 1		His elder brother is Asabudon Ovaisi, Member of Parlament representing				E / Eff Ji Copy @ Delete HV/2507Eallow 12.3r.dd.behoa P	12 yr,old beheads man al gaeda training video hafurb urbenir, strata albadi thrastana Noodibad b	0.111111	0.222222														
an die Wert / Eine Addat. Ander Hande in 3 De 200 g: // Ein Sche g Schen Strifted on teruserbanden prinziptionen. 135 125 g: // Ein Sche g Schen Kellischen in derschen schwarten mit Aussergenzu. 1 13	This Community Processor Speech				Est Stopy Object inStration	beheading																	
m /feft jürge gölen könsätte akandstraus kydekatarin sa akang tet. 0 055					was delivered at Nirmal, Adilabad, Andhra Pradesh on 24 Dec 2012												🗄 🥜 Eét 🕌 Copy 🥥 Delete ((YRgl-co3)	hate speech on indian planning kill/hindus,	0.375	0.25			
															a		Edt H Copy Obiete MEDaddiew/c	akbaruddin owaisi hyderabad mim mla abusing lord r	0	0.25			
	-	0	0) 🧉 😫							• 6	8.40	9 m	152 2-7114	🚯 🍐 📜 🛛	0 0 0 0 0 0	12	4 33333		· [*	2.9	Į

Nex 1ab X	A boanted/ myster wanter x +		File Edit Search View Decoding Language Settings Macro Run Plugies Window ?
🔶 🖲 localhest 'phomysdmin's diphol	lds+prachilitable+der_tokdoerver+18target+8toker+6aad201666002x27238512869x65049M4UR;-16-og.phplds+p=₹	C P-Askcon P ☆ 白 ♣ ★	120H0 5 5 3 2 5 5 0 C 8 5 1 4 1 5 5 1 2 1 4 4 5 5 5 1 2 1 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
nhoMuAdmin	🖬 🚛 Sener, mysaj somponer s 👩 Databara, por la s 🕎 Table, des Juk	X	🗧 Balanzia (Bitanza) (Bitanza) (Bitanzia) (
23009	🗐 Browse 💥 Structure 🔡 SQL 🔍 Search 🐉 Insent 👼 Export 🗃 Import	*) Privileges 🎤 Operations 🕮 Triggers	 Horvmere, kill, aurder, blod, hate, vollence, horrible, scary, homb, specen, qua, firing, kinag, demaid, freedom, pub, best, croel, energ, tipts, hurr, wood, pain, famile, tarteror, a fire, threat, ourflort, appresive, anti, crize, kinfe, scarytive, propo, intergroup, hide, tarpet, kinoryt, procesh, hidotean, pakinstan, india, kashair, bhindustan, lahore, af bash, bash, blogh, bash, brand, blogh, bash, /li>
(Recentitables) .	• Options • · · · · · · · · · · · · · · · · · · ·	des_per des_ppl	 Pranof Mahjan, https://worty. Journal. 1000000000000000000000000000000000000
information_schema	Edit H Copy Oplate 128,xGY17Dc surveillance.sideo.shows.man.charging.four.new.yor.	0 0.0666667	4 Air Force, RAW, PMLS, RATO,TTP,Threek-E-Taliban, Pakistani People Party, Operation Studii Karan, Aryavrat, Conspiracy, Leader, Democracy, Ingalab, Avami, Strike, Whilafet, A
B	Copy Objects 30072XeSkyc .47.6.watched.wideo.eyes.broken.teac.dont.know.eve	0.190476 0.0761905	Party, Leegue, Frotest, Politician, Slopes, Division, Public, Freedent, Secularism, Conservic, Congress, Election, Mitnessed, Tribel, Ballies, supermetal Ymarh, 100, Nowh, Mark Natark, Rolvan, Warferan, Tribute, Soldier, Juan, English, Fishtins, Evening, Leecing, Annual Party, Leecond, Status, Stat
8-3 permanan junera	📋 🥜 Edit 👫 Copy 🥥 Delete 7ukT/sz2/b2o akbaruddin, owaisi, indian. politician, mla, belongs, in	0.0534476 0.041958	j gands, Army, Protocol, Security, Anthem, Threet, Nikes, Border, Shakeedi, Military, Jindaed, Mijah, Sirty Kar, Black Zay, Terror, Kission,
-ig New	📋 🥜 Edit 🙀 Copy 🥥 Dekte BinAMyWya60 trapped teror, mumbai, hotel, attack, india, al. qaeda,	0.0421607 0.0060241	10 Operation.Juli, Frison, Open Fire, Destruction, Halaal, Grave, Sectarian, Genocide, Encounter, Ghadar, Strategy, Satile Field, Mation, Warn-
e y deno	Efek \$4 Copy Objects cGQwTOsh_0 click.watch.part.1.now.http.bit.ly.Tenn77.lightri	0.0538322 0.00598802	1 Dg. Killing, Lepenary, Lapsing, Umians, seapon, Qarr, Umare, Insulting, Uetron, Accident, Umichar, Fallure, Leng, UVI, Vision, Armed
in the designed	Etit H Copy @ Delete CLgLabSPSE	0 0	13 War, Freedom, Jet Carrier, Police men, Slave, Nonour killing, Faith, Ideology, Thunder, Ubeducated, Awarenese, Debate, Foreign, Leaked, Frees, Affair, Economic, Destiny, Flood, Endgan
+ vid_details	E / Edk 14 Copy @ Delete c.Int/CopHz24 terrorist.hindus.killed.christians.india	0.4 0.6	
i∰-ij test	Edit H Copy Oblete dQ/MRU http://websiteprice.us.tawnee.store.rare.photos.http://	8 8	
	🖹 🥖 Edit 🔓 Copy 🥥 Delete d_XXP/XXXIbyg beheading.student.irsq.war.wally.tunny.towards.en	0.111111 0.111111	
	📄 🥒 Edit. 🙀 Copy 🥥 Delete: ETiplisEVTDpo 🛛 assassivation attempt on journalists streets kabul	0 0	
	🖭 🥜 Est 👫 Copy 🥥 Delete GCeCoL/O2wY syria.homs.baba.anno.kataeb.al.tarouk.hee.nyrian	0.0709231 0	
	🗁 🥜 Edit: 🙀 Copy 🥥 Dekite: HCU2HHVe/10 🛛 uks.channel,4.guantaname.handbook.documentary.uk.c.	0 0	
	Efit 🛃 Copy 🥥 Delete H28zVpGgd death,threat,dawahlims,goes,doc.hunting.minor,	0.125 0	
	📋 🥔 Edit 👫 Copy 😄 Delete Indit/2024/bs	0 0	
	Edit 🚽 Edit 👫 Copy 🥥 Delete HM2907Eal/low 12. yr.old. beheads, man.al. gaeda, training video	0.111111 0.222222	
	🗇 🥖 Edit 🕌 Copy 🤤 Delete 10URDpRER 🛛 video edited, show james, folinys, beheading video rel	0.0533901 0.0242710	
	Edt H Copy Oplate ins31skV3M copyright,2008,jhadworksbothways,please.dort.send	0.0066667 0	
	C 2 Edit 34 Copy 2 Delete (2/Rpl-cv3) video uploaded mobile phone.	0 0	
	🗄 🥖 Erik 🙀 Copy 🥥 Delete 1e80xs34x/kc akbaruddin,owaisi,speach,arthapur,randed,jalsa,08,	0.0507012 0.0390625	
	Dekte MGgBuBitto4 mardecdavid harres s militart kordon accent wil	0.0172913 0	· · · · · · · · · · · · · · · · · · ·
	Pl 🖉 Edit. Si Copy. 😩 Debte: nDYhogz7JA	0.0588235 D	Normal teef like length 28/5 lines 13 Let 1 Cel: 1 Sel 2 Decilitedees AM3 P6
🚱 🥝 🥞 🔞	💿 📵 😵 🔛 💽 💷	- P 8 4 6 151	- 😌 🕘 💭 💽 🔮 🚰 😫 😫

Figure 10. Description Tokens and word Lexicon



Figure 11. Detail Page and Result Page

VII. CONCLUSION

In this paper, a classifier (ExClassifier) is proposed to detect the privacy violating hate, extremism and crime producing videos having objectionable and crime content on YouTube. It indicates that the presence of bias features can be used to exploit the hate and extremism detection on YouTube. Many features based on metadata and audio signals along with our manual analysis and visual inspection for each category has been used and a threshold value for each video is calculated. Real world test dataset are analyzed using YouTube APIs. The performance of classifier reveals that certain features like presence of Hate-terms features of the video are important for the accuracy of the proposed approach. The recall rate of the classifier is 84% and the combination of metadata and audio performs better than only the textual metadata such as title and description. If the audio is labeled as extremist video in stage 1 and in stage 2 if the audio contains gunshots

and screaming the video is classified as extremist. Although our experiments show promising results, this research can be extended in future by taking video samples into consideration. The test bed can include videos from other social media websites. The meta-based classifier and the audio based classifiers can be extended by using new features for text and audio to further enhance the results.

References

- [1] Chunneng Huang, T. Fu and H. Chen "Text-based video content classification for online video-sharing sites" in Journal of American society for Information Science and Technology, volume 61. Issue 5, pages 891-906, may 2010.
- [2] Sureka, A., Kumaraguru, P., Goyal, A., & Chhabra, S. (2010). "Mining YouTube to discover extremist videos, users and hidden communities" in *Information retrieval technology* (pp. 13-24). Springer Berlin Heidelberg.
- [3] H. Chen., W.Chung, J.Qin, E.Reid, and M.Sageman, Uncovering the dark web: A case study of jihad on the web. Journal
- [4] K. Filippova, and Keith B. Hall. "Improved video categorization from text metadata and user comments." Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval. ACM, 2011.
- [5] S Mukherjee, P Bhattacharyya "YouCat : Weakly Supervised Youtube Video Categorization System from Meta Data & User Comments using WordNet & Wikipedia" in *Proceedings of COLING 2012*: Technical Papers, pages 1865–1882,COLING 2012, Mumbai, December 2012.
- [6] Tianjun Fu, Ahmed Abbasi, and Hsinchun Chen "A Focused Crawler for Dark Web Forums" in *Journal of the American Society for Information Science and Technology*
- [7] Bermingham, A., Conway, M., McInerney, L., OHare, N., and Smeaton, A.F." Combining social network analysis and sentiment analysis to explore the potential for online radicalization" *IEEE International Conference on Advances in Social Network Analysis and Mining* (Washington, DC, USA, 2009), pp.231236.
- [8] T. Chalothorn, J. Ellman "Using SentiWordNet and Sentiment Analysis for Detecting Radical Content on Web Forums" in 6th Conference on Software Knowledge, Information Management and Application. (SKIMA 2012), 9-11 September 2012, Chengdu University.
- [9] http://www.youtube.com/yt/press/statistics.html
- [10] T. Giannakopoulos, D. Kosmopoulos, A. Aristidou, S. Theodoridis "Violence content classification using audio <u>features</u>" in Advances in Artificial Intelligence 4th Helenic Conference on AI, SETN 2006, Heraklion, Crete, Greece, May 18-20, 2006. Proceedings
- [11] Giannakopoulos, Theodoros, et al. "Audio-visual fusion for detecting violent scenes in videos." Artificial Intelligence: Theories, Models and Applications. Springer Berlin Heidelberg, 2010. 91-100.
- [12] Zou, Xingyu, et al. "Multi-modal based violent movies detection in video sharing sites." Intelligent Science and Intelligent Data Engineering. Springer Berlin Heidelberg, 2013. 347-355.
- [13] Gerosa, Luigi, et al. "Scream and gunshot detection in noisy environments." 15th European Signal Processing Conference (EUSIPCO-07), Sep. 3-7, Poznan, Poland. 2007.
- [14] Pikrakis, Aggelos, Theodoros Giannakopoulos, and Sergios Theodoridis. "Gunshot detection in audio streams from movies by means of dynamic programming and bayesian networks." Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on. IEEE, 2008.
- [15] Vozarikova, Eva, Jozef Juhar, and Anton Cizmar. "Dual Shots Detection." Advances in Electrical and Electronic Engineering 10.4 (2012): 297-302.