YouTube X-Rating Detection with Bahasa-Slang Title Using Query Expansion and Rule Based Approaches

Dewi Wisnu Wardani[1], Salsabila F. Shabihah[2]
[1,2]Department of Informatics, Universitas Sebelas Maret, Indonesia

Article Info

ABSTRACT

The detection of X-rating content on the Internet is still rarely done in Indonesia and the performance of the existing work to detect X-rating content, especially in video is still low. The largest video portal, YouTube, does not yet have automatic X-rating content detection through its content either. Some X-rating content prevention service providers in Indonesia, such as the Internet Positive and Nawala Project, detect X-rating content using the keyword detection method of a web page and then block the web page with DNS filtering. However, that method does not pay attention to using Bahasa-Slang. This work developed Metasearch named Safedio. Safedio aims to detect X-rating content on YouTube content through video titles that contain Bahasa-Slang. Safedio utilizes Query Expansion and Rule-Based approaches. The Query Expansion is a technique to get additional rules in search. In the end, Safedio can detect X-rating content through video titles in both Bahasa and Bahasa-Slang. The average results return with precision 71%, recall 46% and accuracy 72% .

Corresponding Author:

Dewi Wardani,
Departmen of Informatics,
Universitas Sebelas Maret,
Jl. Ir Sutami 36 A, Surakarta 57126 Central java
Email: dww_ok@uns.ac.id

1   INTRODUCTION

Nowadays, the prevention of X-rating content access on the internet is still fairly minimal. This issue is a concern because of the large number of internet users in Indonesia, which is reached 171.17 million people [1]. Some previous work have been concerned in detection X-rating content as well [2] [3] [4] [5]. YouTube is the largest video portal. Still, up to this year 2021, it has not been able to filter X-rating content through automatic content detection. YouTube only provides a restriction mode setting that restricts improper video access where the assessment is obtained from the calculation of inappropriate content flags given by the user. To prevent X-rating content access, Indonesia has an Internet Positive website and Nawala Project that performs DNS filtering of X-rating web URLs using the keyword method. This method still has lack because it only pays attention to the site name and URL.

Related work that discusses the prevention of X-rating content on the internet has been conducted. Several studies regarding the filtering of X-rating words were performed [6] by using text filtering and [7] by using a combination of image detection with IP filtering and text filtering systems. Other studies using the Bayesian approach to detect X-rating content were carried out by [8] on email spam. Bahasa-*Alay* or Bahasa-*Gaul* which can be defined as Slang in Bahasa utilizes the combination of letters, numbers, or symbols that can represent the form of letters [9] [10]. It is mentioned that the main function of using Bahasa-Slang is to hide or keep secrets. So that only certain circles know the contents of the conversation. Based on the statement, the use of Bahasa-Slang on X-rating content will make detection more difficult to pass filtering which still uses standard Bahasa. The varied Bahasa-Slang becomes one of the forms of expansion queries in standard Bahasa that changes shape. In work conducted by [9] [7], one of the variations of Bahasa-Slang uses a combination of letters with numbers and symbols that have a shape resembling a letter. Other work more concern in detecting negative content based on image processing [11] [12] [13] [14] [15] [16]. Therefore, those above existing work which consider in using text-based approach such as text filtering or text combination etc have never considered slang terms. The performance were acceptable, but

they never show in case that text contains slang terms. Meanwhile, slang terms usually are utilized to hide the real intention included in X-rating terms. The appraoch which based of text processing is still promising.

Related work which discussed study Query Expansion and Rules-Based work have been conducted. A few work [17] [18] conduct task on Query Expansion. work that using the Rule-Based method was conducted by [19] be entitled Validating Rule-based Algorithms. Based on those works, it can be concluded that the effort to detect X-rating content on the internet is done through text detection and DNS filtering. YouTube as the largest video portal does not either have automatic detection of X-rating content. Existing text detection methods still have limitations that are only applied to the uses of standard words in the Bahasa. X-rating content that uses non-standard words such as Bahasa-Slang, cannot yet be processed. It makes detection becomes more difficult. Bahasa-Slang has been investigated in several studies which more concerning in language or education [20] [21] [22] [23] [24] [25] [26] [27]. Those above work show that to extend the query intention is necessary. The simple rule-based approach is usefull enough as it has shown good return obtained scores. Rule based is usually better choosen method if the work has small dataset instead of machine learning methods. Therefore, this work more consider in utilizing this method.

In summary, this work creates a filtering system for X-rating content from the YouTube portal by building metasearch that applies the Query Expansion and Rule-Based approach to detect X-rating content through video titles. It is named Safedio. Both methods are applied by taking into the characteristics of the Bahasa-Slang, especially in variations that use a combination of numbers and symbols. Therefore, Safedio metasearch can detect and filter YouTube content that is X-rating in using either standard word titles in Bahasa or Bahasa-Slang.

Several studies related to the detection of X-rating content have been done. A work [7] uses text-matching and image detection for detection. The weakness of this study is that detection is very dependent on the success of detecting skin color. Therefore, if the image is not very clear, it will be difficult to do. Subsequent work using Bayesian as a detection method was conducted by [8] . Bayesian methods provide fairly good results in detecting X-rating content. Further work was conducted [6] using text filtering of X-rating words with Boyer Moore only filters words in Bahasa. In summary, the related work has a lack that is not yet able to overcome Bahasa-Slang in text detection. The work [9] [10] mentioned that one variation of the Bahasa-Slang uses a combination of letters with numbers and symbols. The numbers and symbols used in the Bahasa-Slang are those that have a shape that resembles a letter such for example: 1 = i, 2 = z, 3 = e, 4 = a, 5 = s, etc. This is the basis in work to build a Rule-Based that next is used to detect Bahasa-Slang.

The proposed method by [17] [28] conducted work on Query Expansion to expand queries that were obtained from synonyms. A work using the Rule-Based method was conducted [19] titled Validating Rule-based Algorithms. It aims to build a Rule-Based system that can execute validation algorithms based on a graphical rewriting-based solution approach. Work-related to X-rating content prevention has been carried out using text detection and DNS filtering. Text detection can only detect words in standard Bahasa. Whereas, in Indonesia, there is a non-standard language named Bahasa-Slang. Bahasa-Slang is an expansion of queries from standard Bahasa and is used to hide the true meaning of a word, making it difficult to detect.

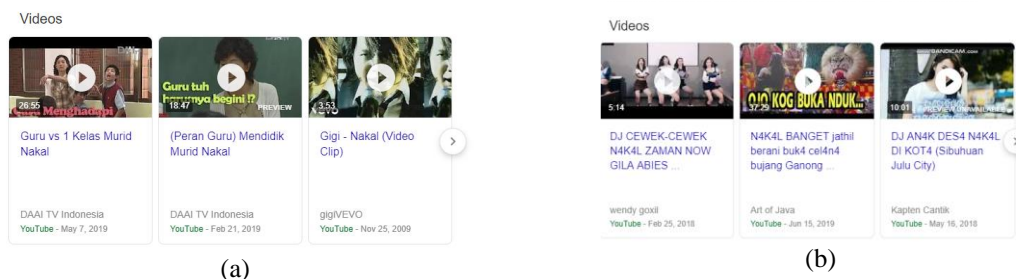(a)   (b)

Figure 1. (a) The result with normal keyword and (b) The result with Bahasa-Slang keyword

The motivation example is shown in Figure 1. Figure 1.a is the result with the keyword "NAKAL" and Figure 1.b is the result with the keyword "N4K4L". The word "N4K4L" is an Bahasa-Slang word. Therefore, there is something it shall be done in handling this issue. In the rest of this document, for the reason of politeness, authors tried to mention less explicit terms and tried to choose not many vulgar terms.

2   METHOD
2.1. Rule Based and Query Expansion Modeling

Bahasa-Slang uses a combination of letters with numbers and symbols [9]. Rule-based built is as follows:

1. Number

Modeling for number categories uses 10 types of numbers found on a standard keyboard, i.e. as follows:

a. "1" = /i/ or /I/   f. "6" = /g/ or /G
b. "2" = /z/ or /Z/   g. "7" = /t/ or /T/
c. "3" = /e/ or /E/   h. "8" = /b/ or /B/
d. "4" = /a/ or /A/   i. "9" = /g/ or /G/
e. "5" = /s/ or /S/   j. "0" = /o/ or /O/

2. Symbol
a. Symbol (~`#%^&*()+=:;'"<>?/|\) are replaced to "",
b. Symbol (-_,.) are replaced to " ",
c. Symbol (@) replaced to "a" or "A",
d. Symbol ($) replaced to "S" or "s"
e. While the special symbol (!) Can be replaced by "i" or "". If the word that has the symbol (!) Which is replaced by "i" becomes a word in the database, then the symbol (!) Is replaced with "i", if it becomes a word that is not in the database, then the symbol (!) Is replaced by " ".

3. Specific words
Rule-Based is built not only focused on variations of the Bahasa-Slang but also certain words that contain X-rating elements. Specific words used in this system are "18+" or "+18" etc where if the word is processed in preprocessing text it will immediately be marked as a X-rating word using the help of Jaccard Similarity. Expansion Query is applied in a syntax-based system to overcome Bahasa-Slang variations that use a combination of letters with numbers and symbols. The word that has been through the parsing phase is then checked whether there is the same word in the Bahasa dictionary and the word porn in the database. If the same word is found, then the word Bahasa-Slang before parsing is expanded in meaning so that it has the same meaning as the word found in the database. As an example :
a. (Bahasa-Slang) M4K4N => (after passing through the parsing phase) MAKAN. In Bahasa's dictionary, there is the word MAKAN, so M4K4N has meaning eat in Bahasa.
b. (Alay Language) S3D1K1T => (after passing through the parsing phase) SEDIKIT. In Bahasa dictionary there is the word SEDIKIT, so S3D1K1T has the meaning a little in Bahasa.

2.2. The Metasearch
This phase is designing the Metasearch program that was built. Following Metasearch Safedio architecture design:
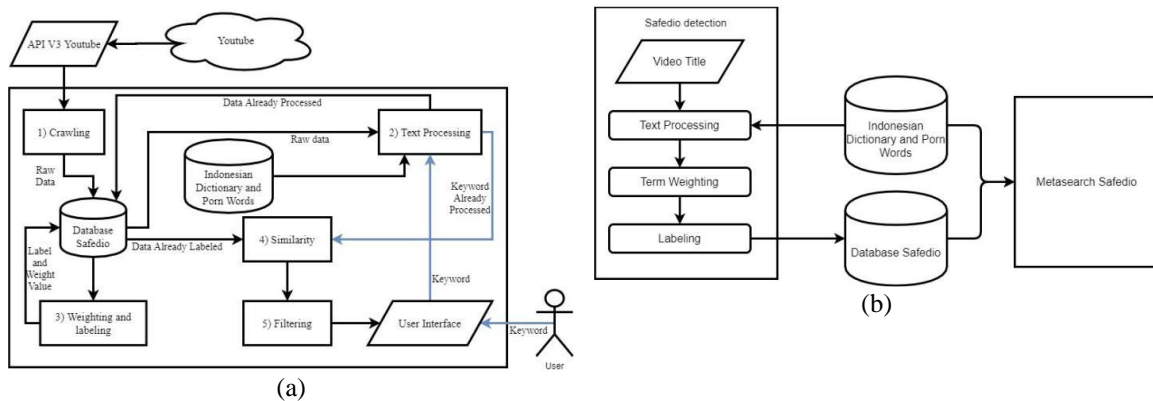

Figure 2. (a) Safedio Metasearch Architecture and (b) Detection Flow in Safedio Metasearch

From Figure 2, Safedio's work generally steps are as follows:
1. Data is collected through a crawler.
2. Data in the form of video titles are processed to get the basic words and parse if there is Bahasa-Slang.
3. The video title is weighted using TF-IDF with the formula proposed by [29]. Labeling video titles using calculated weights and then compared to thresholds obtained from training data and then stored in a database.
4. The user enters a keyword then the keyword is processed in text processing to facilitate searching. Then, finding the similarity between keywords and the term in database. It aims to find the title's intention by users.
5. When the titles that match the user's keywords are found, the program will filter the titles labeled porn so that the titles are not displayed in the search results on the Safedio interface.

Figure 3 describes the detection flow in Metasearch Safedio as follows:
1. Text Processing Video Title
Text processing aims to make the data easier to process and to be structured. Especially for video titles that use slang (Bahasa-Slang). Therefore, they can be processed by the program. For example, the video title

"N1KM4TNY4 B3R®1N®4 DENGAN S3L1NGKUH4N" is done in three phases of Text processing a below:

a. Parsing

Parsing plays the main role in the processing of video titles in programs. The rule-based method is implemented to process the unique characters. It aims to get their complete words. Parsing is done in several categories and is described in Figure 4.

i) Symbol parsing, the earliest phase of parsing in a program is to detect the presence of symbols in the title. The symbols detected are those on the keyboard (~ #% ^ & * () _ + = - `:; '" {} [] | \ /.,> <), Which each symbol is replaced according to Bahasa-Slang rules. After parsing the symbol, the video title then goes through the tokenizing phase, which is to break the sentence of the video title into word for word.

ii) Parsing numbers and letters, this category detects the results of tokenizing that contain a combination of numbers and letters. The check starts with the first word "N1KM4TNY4" and is detected at this phase. The numbers in the words detected in this category are then changed based on the rule-based numbers that have been built in Metasearch Safedio.

iii) Parsing numbers, the next category is the detection of previous parsing results but only for words containing numbers. Alike Parsing that a combination of numbers and letters, the numbers detected are changed based on rule-based rules.

iv) Parsing exclamation points (!), In this category exclamation points have 2 different meanings in work by [9] that can mean the original exclamation point or as a substitute for "i". The results of the previous parsing detected an exclamation point will first be replaced with the letter 'i', then stemming the word. If it is in the Bahasa language in the Bahasa language dictionary or the porn language directory, the exclamation mark in question means the letter 'i'. If not, then the exclamation mark is the original exclamation point and then removed.
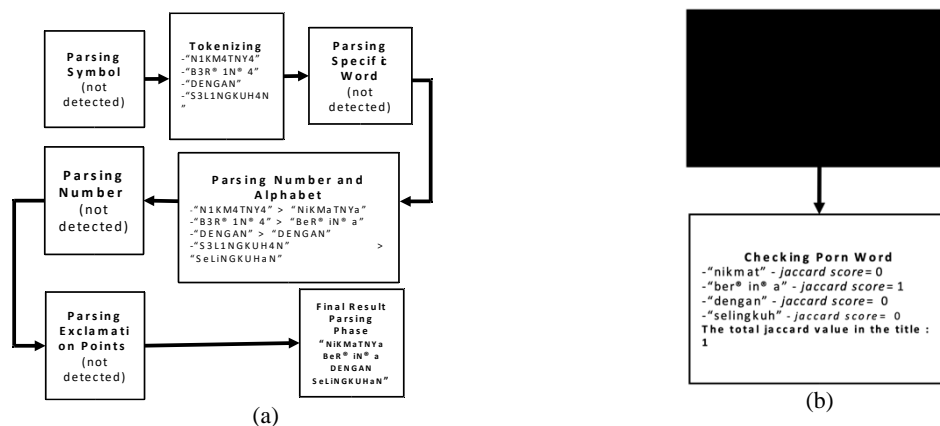

(a)

(b)

Figure 3. (a) Parsing Flow and (b) Stemming Flow

b. Stemming

Stemming is described in Figure 3.b. Words that have been through the parsing phase are then checked whether the word is in a porn word directory or a Bahasa dictionary.

c. Indexing

The outputs of parsing and stemming are video titles that do not contain Bahasa-Slang and the number of words that accept porn. Video titles that have been through the preprocessing text are then indexed to make it easy to search data in a database. Hence, it can save search time and easier in the weight calculation process. Text Preprocessing which is done on the user side is only done in 2 stages, parsing and stemming on entering keywords where there is no need for detection for the word porn.

2. Term Weight

Term weight or this weighting stage uses the TF-IDF weighting method. The terms used in weighting are 72 terms in the porn dictionary as listed in Table 1.

Table 1. List of porn terms

No	Term	No	Term	No	Term	No	Term
1	18	19	Erot®®	37	Kont®®	55	Pel®®
2	Bercin®®	20	Erot®®	38	Lac®®	56	Pen®®
3	Bernaf®®	21	Fu®®	39	Les®®	57	Pep®®
4	Bersetub®®	22	Gigo®®	40	Ma®®	58	Por®®
5	Bira®®	23	Gre®®	41	Masturba®®	59	San®®
6	Bit®®	24	Handj®®	42	Mem®®	60	Se®®
7	Blowj®®	25	Hent®®	43	Mendes®®	61	Sek®®
8	Bok®®	26	Ho®®	44	Menidu®®	62	Selangkang®®

YouTube X-Rating Detection With Bahasa-Slang Title Using Query Expansion and Rule Based Approaches (Dewi Wardani[1], Salsabila Shabihah[2])

154

No	Term	No	Term	No	Term	No	Term
9	Bug®®	27	Hor®®	45	Mes®®	63	Sem®®
10	Cab®®	28	H®®	46	Mont®®	64	S®®
11	Cium®®	29	Jala®®	47	Nen®®	65	Se®®
12	Co®®	30	J®®	48	Ngent®®	66	Telanja®®
13	Colm®®	31	Jilbo®®	49	Nge®®	67	Terangsa®®
14	Cr®®	32	Kelam®®	50	Orgas®®	68	Test®®
15	Da®®	33	Kemalu®®	51	Pant®®	69	Tet®®
16	Des®®	34	Ki®®	52	Pasc®®	70	Tit®®
17	Dub®®	35	ki®si®®	53	Payuda®®	71	Tok®®
18	Ejakula®®	36	Kond®®	54	Pel®®	72	Vagi®®

Then the weight value of each term in the title is summed and averaged it will produce a weight per title. In the training results, this weighting becomes a reference to get a threshold. The threshold weight value is obtained 0.213. The score is obtained as the average weighted results in the training dataset explained in Table 2.

Table 2. The obtained result considering the threshold

Threshold	Number of Titles Above Threshold (Porn)	Number of Titles Below Threshold (Not Porn)	Total of Data
0.213	541	444	985

3. Labeling

The labeling stage is carried out to determine whether the video title is X-rating or not, this stage is done three times. The first uses stemming results, the second is labeling by experts, and the third uses the comparison of weighting results with thresholds.

a. The first labeling that is using the results of detection of X-rating words from the stemming stage is done automatically by the program. If one word is detected X-rating, then the title is the title of a X-rating video. Labeling this stage is used as a reference to select the video title data along with the second labeling in the training dataset to calculate the threshold.

b. The second labeling is manual. Manual labeling is done to label X-rating or not X-rating videos in the title based on the provisions of Law No. 44/2008 on X-rating content. Labeling is used as a reference for selecting video title data together with the first labeling in the training dataset to calculate the threshold.

c. The third labeling uses the results of comparison of the weight value of the video title with the threshold value by the program. If the weight value of the video title has a value equal to or greater than the threshold, the title is labeled X-rating. Conversely, if the weight value of the video title is smaller than the threshold, then the title is labeled not X-rating. This stage is the final labeling used for Metasearch Safedio in selecting the video title that will be raised.

4. Similarity

The purpose of the similarity stage is to find the search results of the keywords entered by users who have done the previous text processing stage with data sources from programs that have passed the stages of text processing, weighting, and labeling.

The similarity in the Safedio system is used to find information needed by users through keywords and compared with data in the Safedio database. The steps taken are as follows:

a. Users search by entering keywords into the system.

b. The keyword entered by the user is done text processing to get the basic words to facilitate the search process.

c. Then do similarity between keywords with words in the index data. Title documents that contain the same index term as the user's input keywords are then collected by the program.

5. Ranking

The ranking is based on the results of labeling stage three where the weight values of each title are sorted. The top-ranked is measured from the matching score between the title and the terms of the user's keywords and that obtained the lowest weight.

Ranking on Metasearch Safedio is not done specifically. The ranking uses the similarity of results from the user's keywords and sees the weight of the sequential titles from lowest to highest. The first video title is the title that has all the keyword elements of the user and has the lowest weight. The next video title has at least 1 element of the keyword enter the user and has the lowest weight to the highest.

6. Filtering

Filtering is done to sort out which videos contain X-rating elements and which videos do not contain X-rating elements. This is based on the third labeling process. X-rating videos have video titles with weights above the threshold value, while videos that do not have X-rating content have video titles with weights below the threshold value.

3   RESULTS AND DISCUSSION

3.1. Dataset

Data were collected using a crawler built on the YouTube v3 API to search data titles, and video links by entering keywords, the amount of data, and search categories by date, rating, relevance to keywords, titles, and a number of views.

Training datasets were collected using a crawler from Safedio. The training dataset is obtained by using keywords that are X-rating and not X-rating. For keywords containing X-rating elements obtained from the bad-word list on Google downloaded through the website www.freewebheaders.com update on July 30. 2018, which was previously translated into Bahasa, while keywords that do not contain X-rating elements are used random keywords on political, religious themes, etc. It takes time to label the data from the time its collected, until it is ready to be processed. It has been checked that the data is not much different with the recent data. The keywords that contain X-rating elements are inputted either in the Bahasa language standard or transformed into Bahasa-Slang variations that use a combination of letters with numbers and symbols.

Dataset Testing is divided into 3 categories of data-sets :

1. Random dataset, the testing data is obtained randomly from the website www.ytgenerator.com which does not see any limitations on the work by pressing the "Generate Video" button for 100 data titles and video links. As the beginning work, it is used little data first. Obtained 2 titles containing X-rating and 98 titles not containing X-rating from manual assessment.

2. Non-Bahasa-Slang dataset, video title data that is searched based on 50 keywords from the word porn which is found in the index results directory which has the most frequent appearance of training data with the condition that the video title labeled porn on the first labeling and the second labeling. As the beginning of work, the number of keywords is just an example, as it is obtained from terms that frequently used. This number is The keyword does not contain elements of the Bahasa-Slang. Searching this dataset through the crawler has 46 porn titles and 54 non-porn titles from manual assessment.

3. Bahasa-Slang dataset, video title data that is searched using the same keywords as the Non-Bahasa-Slang dataset search. The keyword is then converted into vocal letters such as variations in the Bahasa-Slang. Searching this dataset via the crawler has 53 porn titles and 47 non-porn titles from the manual assessment. The data collected is used as a training dataset and testing dataset with details in Table 3.

Table 3.   The dataset

No	Dataset Type	Amount of Data Porn	Amount of Data Non-Porn	Amount of All Data
1.	Training Dataset	565	420	985
2.	Random Dataset	2	98	100
3.	Non Bahasa-Slang Dataset	46	54	100
4.	Bahasa-Slang Dataset	53	47	100

3.2. The Results

Testing is done by performing on 3 categories of data sets. They are the random dataset, non-Bahasa-Slang dataset, and Bahasa-Slang dataset. The results of the label by the program if it has a value of 1 then the title contains X-rating elements, if it is 0 then the title does not contain X-rating elements. The true/false column shows the same value between the program label and the manual label. Weight is obtained from TF-IDF computation. Program Label is the obtained label based on this proposed approach. Manual label is the label from what its said expert. The label True or False is based of the factual or manual label.

Table 4. The example of test result over Random Dataset

No	Title	Title of Text Processing Result	Porn Word	Weight	Program Label	True / False
1	"Clint, Give It To Me" Black Panther Running With Gauntlet - AVENGERS:ENDGAME (2019) 4K Movie Clip	Clint Give It To Me Black Panther Running With Gauntlet AVENGERS:ENDGAME 2019 4K Movie Clip	-	0	0	true
2	"We Are Venom" Scene - VENOM (2018) 4K Movie Clip	we are Venom Scene VENOM 2018 4K Movie Clip	-	0	0	true
3	"What are your weaknesses?" Job Interview Question / 9 Great Answers!	What are your weaknesses Job Interview Question / g Great Answers!	-	0	0	true
4	?I Love Led Zeppelin? Scene - Spider-Man: Far From Home (2019) 4K Movie Clip	i Love Led Zeppelin Scene Spider Man: Far From Home 2019 4K Movie Clip	-	0	0	true
5	01 Introduction. The Land of the Bible: Location & Land Bridge	o Introduction The Land of the Bible: Location Land Bridge	-	0	0	true

In the random dataset test results from Table 4, no titles were labeled porn. A significant comparison between porn and non-porn titles in random dataset testing is because the data obtained are truly random and do not follow any rules so that the titles found can be Bahasa and foreign languages.

Table 5. The example of test result over Non Bahasa-Slang Dataset

No	Title	Title of Text Processing Result	Porn Word	Weight	Program Label	True / False
1	Cara berhubungan badan yang benar Chentub Tutorial	cara hubung badan yang benar Chentub tutorial	-	0	0	false
2	Belajar w®k w®k alias ng®®e sama kaka.. || buat para pasc®l	ajar w®k w®k alias ng®®e sama kaka buat para pasc®l	Ng®®e pa®c®l	0.0556	1	true
3	Anime l®s®i citrus episode 4 makin asik lihatnya	Anime l®s®i citrus episode a makin asik lihat	L®s®i	0.0278	0	false
4	Bigo Live cewek s®®y ngemut pisang - Girls love Bl®wj®b banana	Bigo Live cewek s®®y ngemut pisang Girls love bl®wj®b banana	S®®y bl®wj®b	0.0423	1	true
5	10 ciri wanita yang be®naf®u tinggi diranjang	10 ciri wanita yang be®naf®u tinggi ranjang	Be®naf®u	0.0278	0	true

The example result over the non-Bahasa-Slang dataset has shown in Table 5. Meanwhile, Table 6 shows the example result over Bahasa-Slang Dataset.

Table 6. The example of test result over Bahasa-Slang Dataset

No	Title	Title of Text Processing Result	Porn Word	Weight	Program Label	True / False
1	Oh god he's so ®0t (read desc if you are italian)	Oh god hes so h®t read desc if you are italian	h®t	0.0133	0	true
2	®0t Aunty on Live Cam Live Show TikTok Girls Live Bigo Live Cam 2019	h®t Aunty on Live cam Live Show TikTok Girls Live Bigo Live cam 2019	h®t	0.0133	0	false
3	Jhanvi Kapoor With BHABHEE Malaika Arora Making A ®0T-DU0 Togethar Snapped Wid Each Other For Gym	Jhanvi Kapoor With BHABHEE Malaika Arora making a h®t DU0 Togethar Snapped Wid Each Other For Gym	h®t	0.0133	0	true
4	®0t G0ss!ps..Nana Aba Anamoah R?v?@ls $ecr?ts about Frema Ashkar..Shatta Wale f!r?s Obour Heavily.	h®t G0ss!ps Nana aba Anamoah r v als secr ts about Frema Ashkar Shatta Wale f!r s Obour Heavily	h®t	0.0133	0	true
5	Avika Gor Looks So ®0T in Red Dress || Avika Gor || Raju Gari Gadhi 3 Pre Release Press Meet || MB	Avika Gor Looks So h®t in Red Dress Avika Gor Raju gari Gadhi e Pre Release Press Meet MB	h®t	0.0133	0	true

Table 7. The summary of the testing results over 100 dataset

Random Dataset			Non Bahasa-Slang Dataset			Bahasa-Slang Dataset		
	Porn	Not Porn		Porn	Not Porn		Porn	Not Porn
Porn	0	0	Porn	18	8	Porn	14	6
Not Porn	2	98	Not Porn	28	46	Not Porn	39	41

From table 7 can be summarized the accuracy over Random, Non-Bahasa-Slang, and Bahasa-Slang data sets are 98%, 64%, and 55% respectively. Therefore, the average obtained accuracy is 72%. It is a pretty good score. Weight is obtained from TF-IDF computation.

Table 8. The example result using keyword "Goyang"

No	Title	Weight	Manual Label	Program Label
1	goyang h®®	0.0138	0	0
2	Artis bigo m0®®0k goyang r4nj4ng kelihatan anu nya	0.0199	1	0
3	Awek bigo goyang s®®si	0.0199	1	0
4	Tik Tok Goyang P4®®4t Sampe Bawah Terbaru Bikin Ngilu	0.0235	1	0
5	DJ TIK TOK atas bawah goyang goyang me®®esah Remix 2019 terbaru	0.0260	1	0
6	Goyang sambil t3l4®®4ng	0.0260	1	0
7	Tiktok jilbab goyang ®®t | jilbab m®®tok	0.0338	1	1
8	2 GOBAS GOYANG KELIHATAN PA®®DARA MO®®OK	0.0446	1	1
9	BIGO LIVE H®® | TANTE CANTIK SE®® BERGOYANG | COL®®K	0.0593	1	1
10	Bigo live ®®t cewek mulus & se®® sem®® mo®®ok goyang er®®is	0.1028	1	1

Table 9. The Precision and Recall on 100 key queries over combined dataset

No	Keyword	TruePositive	True Negative	Amount of Porn Label Manually	Amount of Data	Precision	Recall
1	Bigo	9	0	19	20	0.45	0.473
2	Live	10	1	20	22	0.5	0.5
3	Goyang	4	1	9	10	0.5	0.4444
4	Pasangan	0	1	0	2	0.5	undefined
5	Cewek	5	2	11	13	0.5384	0.4545
6	Video	2	6	5	13	0.6154	0.4
7	Film	2	1	4	6	0.5	0.5
8	Movie	1	3	3	6	0.6667	0.3333
9	Mulus	2	0	5	5	0.4	0.4
...	...	...	...	...	...	...	...
100	Mandi	1	1	1	2	1	1

To calculate the Precision and Recall, other experiments have been performed. They are done by input a keyword into the Metasearch. The example of an experiment is shown in Table 8. After 100 times experiment, the results are shown in Table 9. In summary, the average Precision and Recall are 71% and 46% respectively. Based on the calculation results, inconsistent results are shown in the low recall value. A low recall value indicates that there are many titles containing the word porn but the contents of the video do not contain porn.

A small experiment was also conducted to find out whether the title in Bahasa-Slang escaped the selection of X-rating content on YouTube. This experiment was carried out 5 times the search without logging in to the YouTube portal. Examples of the results obtained are shown in Figure 4.

From Figure 4.b, the title with Bahasa-Slang still passes the selection by YouTube. Metasearch Safedio system can overcome this problem. So, that both video titles using Bahasa and Bahasa-Slang can be detected by Safedio.

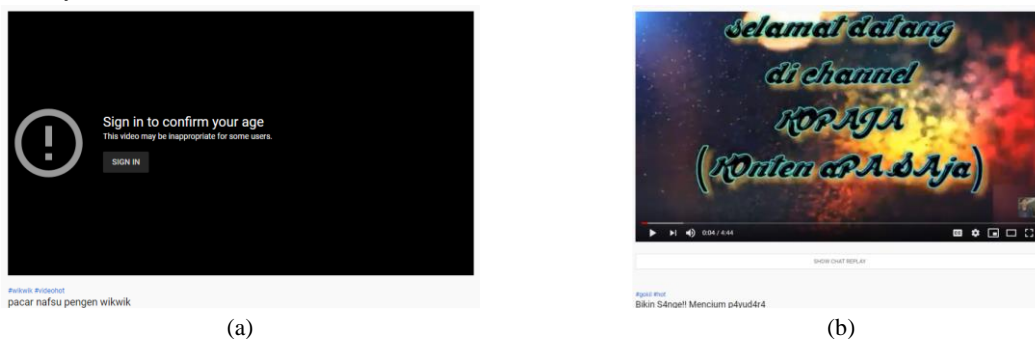(a)   (b)

Figure 4. (a) Videos with Titles that Have Porn Words in Bahasa, (b) Videos with Titles that Have Bahasa-Slang Porn Words in

On the whole test and evaluation results obtained, the Safedio Metasearch has quite good results. Detection of X-rating content with video titles containing Bahasa-Slang which previously could not be filtered by YouTube can be detected using Metasearch Safedio. So that the problem regarding the use of the Bahasa-Slang in Bahasa-language texts on the YouTube video title can be resolved quite well using the Rule-Based and Query Expansion modelling implemented in the Safedio Metasearch. But there are still some things that become limitations in this study. Among others are :

1. There are words in Bahasa, that are not standard and are not separated by spaces so it is difficult to do tokenizing.
2. There are neutral words that become ambiguous when combined with other words such as the words "gituan" in the title "MASA GITUAN D BUS !!! WANITA YANG BAIK HATI".
3. Not yet able to overcome the word porn that uses a combination of neutral words like "ce®e pa®®s".

4   CONCLUSION

This work has succeeded in building a metasearch engine named Safedio. It can detect X-rating content on the YouTube portal through video titles with quite good results with a precision level of 71%, a recall of 46%, and an accuracy of 72%. The limitation of this work shall be overcome such as that there are Bahasa words that are not standard and are not separated by spaces so it is difficult to do tokenizing. Also to determine neutral words that become ambiguous when combined with other words. The near futue work that shall be done are to add rules to Rule-Based to deal with symbols that don't exist on the keyboard and

emoticons, to utilize the other better similarities and using other methods of detecting X-rating content to get better results and to detecting X-rating content other than through titles, such as comments, tags, descriptions, etc.

ACKNOWLEDGEMENTS

5   REFERENCES

[1]   APJII, *PENETRASI & PROFIL PERILAKU PENGGUNA INTERNET INDONESIA SURVEI 2018*. 2019. Accessed: Oct. 30. 2019. [Online]. Available: https://apjii.or.id/survei2018s/download/SPt3b2klhjHziyMBGOq0x4vA8fsDoL

[2]   B. B. Pratamawaty, P. Limilia, and P. Prihandini, "Young people's perception of internet X-rating content: Case of junior high school students' in West Java Indonesia," *International Journal of Psychosocial Rehabilitation*, vol. 24, no. 1, pp. 492–501, 2020.

[3]   Z. Ardi, K. Viola, and I. Sukmawati, "An Analysis of Internet Abuses Impact on Children's Moral Development," *JPPI (Jurnal Penelitian Pendidikan Indonesia)*, vol. 4, no. 1, pp. 44–50. 2018.

[4]   R. Panuju, "PUBLIC NON-COMPLIANCE WITH PROHIBITION OF ACCESSING X-rating content ON THE INTERNET: A CASE IN INDONESIA," *Journal of Critical Reviews*, vol. 7, no. 19, pp. 1274–1285, 2020.

[5]   R. Puspita and D. Rohedi, "The impact of internet use for students," in *IOP Conference Series: Materials Science and Engineering*, 2018, vol. 306, no. 1, p. 012106.

[6]   E. M. S. Sulhan, "Text Filtering Kata Porno dengan Metode Boyer Moore pada Aplikasi Elearning Berbasis Cms di Sdn," *Bimasakti : Jurnal Riset Mahasiswa FTI UNIKAMA*, 2014, Accessed: Dec. 04, 2019. [Online]. Available: https://www.neliti.com/publications/182947/text-filtering-kata-porno-dengan-metode-boyer-moore-pada-aplikasi-e learning-berb

[7]   D. A. W. Nurhayati, "MORPHOLOGICAL AND MORPHOPHONEMIC PROCESS OF ALAY VARIATION," *LINGUA: Journal of Language, Literature and Teaching*, vol. 12, no. 1, pp. 59–70. 2015, doi: 10.30957/lingua.v12i1.71.

[8]   D. Wardani, R. Siwi, B. Harjito, and M. Marshallia, "Using Metadata in Detection Spam Email with X-rating content Content," in *2018 International Conference on Electrical Engineering and Computer Science (ICECOS)*, 2018, pp. 213–218.

[9]   N. D. Kelana, "FENOMENA BAHASA ALAY: PROSES PEMBENTUKAN DAN IMPLIKASINYA TERHADAP PERKEMBANGAN BAHASA INDONESIA," Thesis, Universitas Diponegoro, 2011. Accessed: Dec. 04, 2019. [Online]. Available: https://fib.undip.ac.id/digilib/home/fib.undip.ac.id/files/e_book/Fenomena%20Bahasa%20Alay%20(Proses%20Pe mbentukan%20Kata%20dan%20Implikasin.pdf

[10]   S. S. Andriani and I. D. P. Wijana, "BAHASA ALAY," Doctoral dissertation, Universitas Gadjah Mada, 2016. Accessed: Dec. 04, 2019. [Online]. Available: http://etd.repository.ugm.ac.id/index.php?mod=penelitian_detail&sub=PenelitianDetail&act=view&typ=html&bu ku_id=94623&obyek_id=4

[11]   M. B. Garcia, T. F. Revano, B. G. M. Habal, J. O. Contreras, and J. B. R. Enriquez, "A pornographic image and video filtering application using optimized nudity recognition and detection algorithm," in *2018 IEEE 10th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, 2018, pp. 1–5.

[12]   E. Yiallourou, R. Demetriou, and A. Lanitis, "On the detection of images containing child-pornographic material," in *2017 24th International Conference on Telecommunications (ICT)*, 2017, pp. 1–5.

[13]   H. Li, X. Guo, H. Wang, B. Yu, and P. Zhao, "Application of Computational Complexity Theory to Pornographic Detection in Live Based on Frame Sampling," in *2020 IEEE 6th International Conference on Computer and Communications (ICCC)*, 2020. pp. 1334–1338.

[14]   J. Mallmann, A. O. Santin, E. K. Viegas, R. R. dos Santos, and J. Geremias, "PPCensor: Architecture for real-time X-rating content detection in video streaming," *Future Generation Computer Systems*, vol. 112, pp. 945–955, 2020.

[15]   A. Tabone, A. Bonnici, S. Cristina, R. A. Farrugia, and K. P. Camilleri, "Private Body Part Detection using Deep Learning.," in *ICPRAM*, 2020. pp. 205–211.

[16] C. Tian, X. Zhang, W. Wei, and X. Gao, "Color pornographic image detection based on color-saliency preserved mixture deformable part model," *Multimedia Tools and Applications*, vol. 77, no. 6, pp. 6629–6645, 2018.

[17] C. Carpineto and G. Romano, "A Survey of Automatic Query Expansion in Information Retrieval," *ACM Comput. Surv.*, vol. 44, no. 1, pp. 1–50. 2012, doi: 10.1145/2071389.2071390.

[18] B. He and I. Ounis, "Studying Query Expansion Effectiveness," in *Advances in Information Retrieval*, Berlin, Heidelberg, 2009, pp. 611–619. doi: 10.1007/978-3-642-00958-7_57.

[19] L. Lengyel, "Validating Rule-based Algorithms," *Acta Polytechnica Hungarica*, vol. 12, no. 4, p. 17, 2015.

[20] S. Ekalestari, "Pengaruh Penggunaan Bahasa Alay Terhadap Penggunaan Bahasa Indonesia yang Baik dan Benar," 2017.

[21] A. Wicaksono and S. Supriyono, "TINJAUAN SOSIOLINGUISTIK BAHASA ALAY DALAM KONS^{TEL}ASI KEBAHASAAN," *Ksatra: Jurnal Kajian Bahasa dan Sastra*, vol. 3, no. 1, pp. 19–28, 2021.

[22] A. Lutfiatuna, A. Novitasari, and A. Helfiyana, "Bahasa Alay Pada Chating Di Medsos Remaja Millenial (Bahasa Alay Vs Remaja Millenial)," in *Prosiding Seminar Nasional Bahasa dan Sastra Indonesia (SENASBASA)*, 2018, vol. 2, no. 2.

[23] K. Inoue, "文字でみる若者言葉―インドネシアの Bahasa Alay を題材として―," *Ayumi: Jurnal Budaya, Bahasa dan Sastra*, vol. 4, no. 1, 2017.

[24] L. Laelasari, L. Oktavia, and I. Mustika, "Pengaruh bahasa alay terhadap penggunaan bahasa indonesia di kalangan mahasiswa ikip siliwangi," *Parole (Jurnal Pendidikan Bahasa dan Sastra Indonesia)*, vol. 1, no. 5, pp. 675–680. 2018.

[25] S. Abid, "KEBERADAAN BAHASA ALAY DALAM PERKEMBANGAN BAHASA INDONESIA," *Jurnal Perspektif Pendidikan*, vol. 11, no. 2, pp. 22–29, 2017.

[26] P. DEAMA, S. Sayuti, and I. Romi, ""Penggunaan Bahasa Alay Remaja Pada Status Media Sosial Facebook: Tinjauan Sosiolinguistik," PhD Thesis, UNIVERSITAS BUNG HATTA, 2020.

[27] S. Agustina, D. Indihadi, and H. Hodidjah, "ANALISIS PENGGUNAAN BAHASA ALAY DALAM KOSAKATA BAHASA INDONESIA SISWA SEKOLAH DASAR," *PEDADIDAKTIKA: Jurnal Ilmiah Pendidikan Guru Sekolah Dasar*, vol. 2, no. 1, pp. 1–8.

[28] Z. Abedjan and F. Naumann, "Synonym Analysis for Predicate Expansion," in *The Semantic Web: Semantics and Big Data*, vol. 7882, P. Cimiano, O. Corcho, V. Presutti, L. Hollink, and S. Rudolph, Eds. Springer Berlin Heidelberg, 2013, pp. 140–154. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-38288-8_10

[29] C. Manning and H. Schutze, *Foundations of statistical natural language processing*. MIT press, 1999.