CAHIERS MAGELLANES-NS
Volume 06 Issue 2
2024

ISSN:1624-1940
DOI 10.6084/m9.figshare.2632599
http://magellanes.com/

NLP AND COMPUTER VISION FOR DETECTING ILLICIT ACTIVITIES ACROSS TWITTER AND EXTERNAL LINKS

Vagulagadda Shireesha[1], P V Ramana Murthy [2]
[1]PG Scholar , Department of Computer Science andEngineering , Malla Reddy Engineering College.,Hyderabad
[2]Associate Professor , Department of Computer Science andEngineering , Malla Reddy Engineering College,Hyderabad

Abstract:
Human trafficking is an issue all over the world and dehumanizes millions of people. Right now, trade networks spread this crime on the web with coded messages to promote such illegal businesses. Thus, since there are already limited resources in the law enforcement system, it becomes paramount to automatically detect messages that may be related to the crime, and which might also lead to further investigations. With the aid of natural language processing, this work groups tweets that could promote these illegal services and exploit minors. Images and URLs contained in such suspicious messages are further processed and are sorted out according to gender and age group, which can detect photographs taken of persons under 14 years of age. The first step involves mining tweets in real time containing hashtags related to minors. The key step is to preprocess the tweets so as to remove background noise and misspelled words after which the tweets can be classified to suspicious or non-suspicious. Face and torso geometrical features are then selected using the Haar model. The use of SVM and CNN allows for such identification concerning the torso and its proportionality with relation to the head, even in cases where the face details are undetectable. When torso features only are used, the SVM method performs better than CNN.

Keywords: Illicit Messages, NLP , CNN, Twitter, SVM, ML, Deep Learning

INTRODUCTION:

Once purely original websites were established for reading, in that nobody could really interact with the web. There was one radical change from the inception of web 2.0 after the user was no longer a spectator; they became an active person in social networking such as Facebook, Twitter, and Instagram. To my great worry has also come the door to malign business such as human trafficking, wherein countries of Latin America among others have been at the top for the smuggling of human beings above all children and teenagers aged 14 years and below. What is vital to note is that the average age of sexuality consent is 14 years across Latin American countries with the preference that if minors are somehow made to attend illicit sexual services they shall be regarded as direct victims of the crime of human trafficking. Currently, one can see on Twitter ads about escort services or similar services that promote young girls for consumers. Girls in these businesses are more likely to be abused and exploited sexually, psychologically, and physically.

Many criminal organizations have turned to advertising these "sexual services" on social networks under the cover of various seemingly innocent terms, such as chicken soup-a euphemism for child pornography. Through the internet, criminal organizations can expand their illicit activities, making it possible for covert advertising and messaging to be employed to promote illegal services for the exploitation of unsuspecting minors. There have been prior works on modeling tweet and image classifiers designed to filter out illicit messages independently by employing either natural language processing methods or computer vision techniques. Nevertheless, the current approach exhibits disparate treatment of text and imagery across NLP and Computer Vision techniques. In this paper, the authors concentrate on the web-based advertisement and analyze them for automatic detection of suspected messages. This work employed 10,000 ads that were manually annotated for this purpose. The work focuses on the classification of ads with text and images combined together for analysis. The Human Trafficking Deep Network is a multimodal deep learning model for which they report an F1 value of 75.3%, a recall of 70.9%.

In contrast, contemporary image classification models only utilize facial information, disregarding the fact that many of the images will have the face blurred. The authors in used computer vision algorithms to establish age classification accuracy, up to 86.64%. Another classification method includes SVM and CNN used for predicting the gender of a person. There are no known works that consider the upper torso in images when classifying age groups. The present work consists of two phases. In the first phase, NLP techniques are used for the identification of Twitter messages promoting illicit services offered by minors. In the second phase, pictures are extracted from the suspected websites so as to have processing and gender recognition into two groups of: over 14 years and under or equal to 14 years. Facial features and torso characteristics intent recognition. Many of them clearly have blurry pixels.

System Architecture:

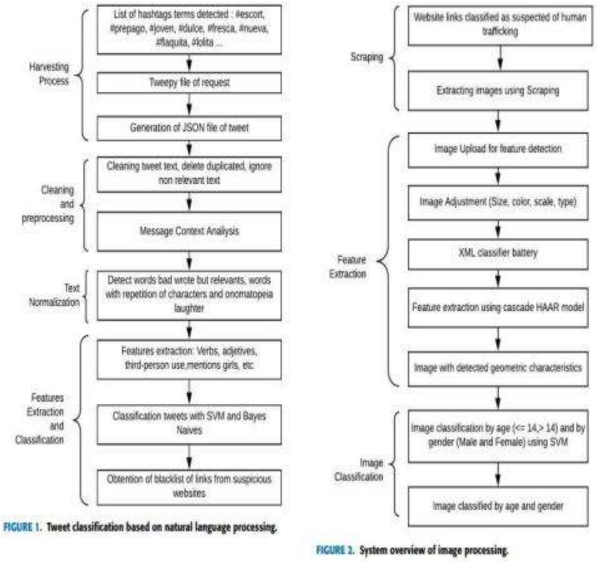

Harvesting Process
List of hashtags terms detected : #escort, #prepago, #joven, #dulce, #fresca, #nueva, #flaquita, #lolita ...
Tweepy file of request
Generation of JSON file of tweet
Cleaning and preprocessing
Cleaning tweet text, delete duplicated, ignore non relevant text
Message Context Analysis
Text Normalization
Detect words bad wrote but relevants, words with repetition of characters and onomatopeia laughter
Features Extraction and Classification
Features extraction: Verbs, adjetives, third-person use,mentions girls, etc
Classification tweets with SVM and Bayes Naives
Obtention of blacklist of links from suspicious websites
Scraping
Website links classified as suspected of human trafficking
Extracting images using Scraping
Feature Extraction
Image Upload for feature detection
Image Adjustment (Size, color, scale, type)
XML classifier battery
Feature extraction using cascade HAAR model
Image with detected geometric characteristics
Image Classification
Image classification by age (<= 14,> 14) and by gender (Male and Female) using SVM
Image classified by age and gender

FIGURE 1. Tweet classification based on natural language processing.

FIGURE 2. System overview of image processing.

Fig : Architecture

Problem Statement :

This inability to gather feedback frustrates the management of each and every unmanned restaurant to obtain general information about customer experiences with their concept and food. Existing rating systems, Google and TripAdvisor, remediate, to some extent, this problem since they only cover a little of what customers have to say about their experience. These rating systems are used only by a few of the customers, by those customers that rate the restaurant, more often than not, upon their own discretion on platforms totally independent of one another. It is mostly directed to those customers who feel very strongly one way or another-positive or negative-about their visit.

Model Proposed:

ISSN:1624-1940
DOI 10.6084/m9.figshare.2632599
http://magellanes.com/

It is absolutely essential for any group of customers to provide motivation for rating; this paper develops a method of assessing the restaurant by requiring every customer to provide one following his or her visit, thereby constantly trying to increase the counts. That system will specifically be fit for unmanned restaurants via the scoring system based upon facial expression recognition using pre-trained Convolution Neural Networks (CNN). The customer is allowed to use a photograph of his or her face-as a score-indicating respective feeling by taking or capturing. While not the same amount of information and not at all that each one experiences with this text-based rating system was collected, this simple and straightforward rating system should reflect more opinions about customer experiences with the restaurant concept.

Methodology:

Detecting possible illicit messages using natural language processing (NLP) algorithms involves the application of computational techniques to understand, interpret, and generate human language content in order to identify and flag messages that may be associated with illegal activities. This process typically includes several steps, such as data collection, preprocessing, feature extraction, model training, and evaluation.

Data collection involves gathering a large corpus of text data that includes both legitimate and illicit messages. Preprocessing steps may include tokenization, stemming, lemmatization, and the removal of stop words to prepare the text for analysis. Feature extraction involves converting the preprocessed text into a numerical format that can be understood by machine learning algorithms, such as using bag-of-words, TF-IDF, or word embeddings.

Model training is a critical step where machine learning algorithms, such as Naive Bayes, Support Vector Machines, or deep learning models like Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs), are used to learn patterns from the labeled dataset. These models are trained to distinguish between benign and illicit messages based on the features extracted from the text.

Once the model is trained, it can be used to predict whether new, unseen messages are likely to be illicit. The evaluation step involves testing the model's performance using a separate dataset and metrics such as accuracy, precision, recall, and F1-score to ensure that the model is effective in identifying illicit messages without flagging too many legitimate ones as false positives.

To enhance the detection capabilities, NLP algorithms can also incorporate sentiment analysis, topic modeling, and context analysis to better understand the intent and content of the messages. Data Collection: Gather a diverse dataset of messages, including both legitimate and illicit examples. This data can come from various sources such as social media, chat logs, emails, or other communication platforms**.**

NLP PROCESS

Provides tools and methods to facilitate the cleansing of data prior to analysis. Basic steps in this stage include tokenization, ontology development, stemming and lemmatization, removal of non-informative terms:

Tokenization: Breaking down the text into words or tokens.

Normalization: This includes converting to lower case, removing punctuation while correcting spelling by using any inbuilt methods with NLTK.

Stemming vs. Lemmatization: Reduction of words to their base or root form.

Stopword removal: Removal of common words that may not carry significant meaning in the message.

Feature Extraction: Convert the cleaned data into a format suitable for machine learning models. Many common ways to do this include:

Bag of Words: The categorical representation of the words with consideration to grammar or order.

TF-IDF: Weighting relative frequency of occurrences of words about other occurrences for the same words across the entire corpus.

Word Embedding: Use of pre-trained models such as Word2Vec or GloVe in representing the words in a continuous vector space so as to capture semantic meanings.

Model Training: Apply a machine learning model that would classify the messages as illegal or legal. The very common algorithms users can put into this task include the following:

Support Vector Machine: A classifier that finds hyperplanes that best divide the data into classes.

Deep Learning Models; Convolution Neural Networks (CNN), which are capable of capturing more complex data patterns.

Model Evaluation: Testing the model with separate validation datasets with accurate metrics like accuracy, precision, recall, F1-score, and confusion matrix. If this evaluation finds problematic areas, modify the model and retrain it to improve performance. More sophisticated machine learning approaches may combine multiple models or re-purpose pre-trained models to boost overall detection accuracy**.**

Results

The author has used this approach in the current work to look for the human trafficking phenomenon in text messages in social media using machine learning algorithms SVM and Naïve Bayes. In this paper the author first crawl twitter by using words like Lolita, escort and many more and then the extracted tweets will go for cleaning by removing special symbols and stop words(words like the, where, and, an are etc.) and then the tweets will be analyse for extracting words such as VERBS and ADJECTIVE and these words may contain important subjects or suspicious words used by HUMAN TRAFFICERS (suspicious words can be chicken soup, girls, penguin and many more. Clean up and analyze these tweets, and pass them on to SVM and Naïve Bayes classifier which classifies words as suspicious or not.

So if that particular tweet contains any suspicious word, then the website for that particular tweet will be searched for images and an SVM HAARCASCADE classifier will be used for each image in order

ISSN:1624-1940
DOI 10.6084/m9.figshare.2632599
http://magellanes.com/

to detect a face from that image while the same algorithm will be employed to detect upper body and these two resultant images will be given into the Convolution Neural Networks classifier which will detect or predict AGE and GEND. In this paper, we detect the male and female genders and predict age into two classes: 14 years and below or over 14 years.

Note that for the implementation of this project, your computer must have an active internet connection since the module 'Crawl Twitter' automatically crawls twitter and the module 'Scrape Images from the websites' which scrape images from the provided website addresses through appropriate API.

The modules in the project are:

1) Online Crawl Twitter: In this module, we will provide a HASHTAG and the application will retrieve all the tweets that related to the provided hashtag from twitter by the help of TWEEPY API.

2) Offline Upload Twitter Dataset: In case one doesn't want to crawl twitter in this module, an existing twitter dataset will be uploaded instead.

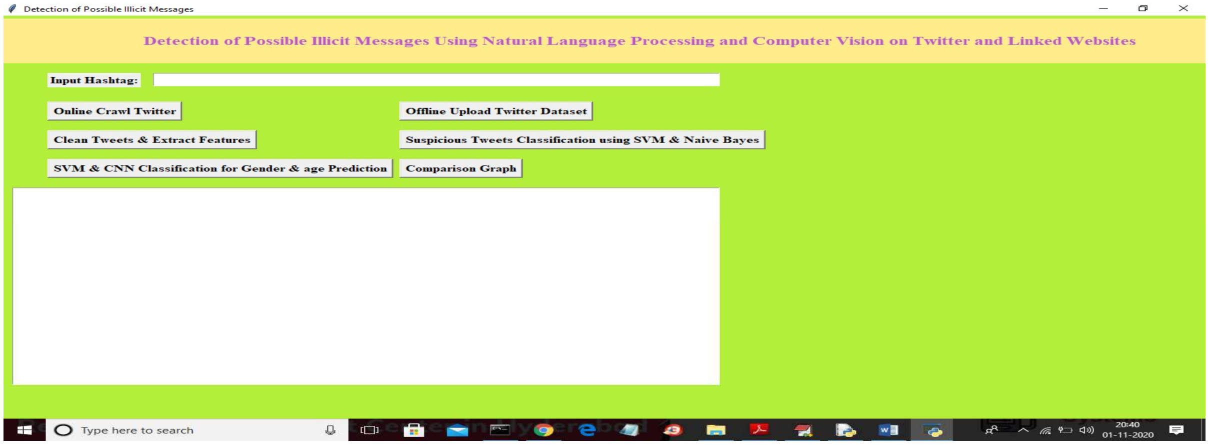
3) Clean Tweets & Extract Features: This module processes every one of the tweets. It wipes off special symbols and stop words, extracts VERBS and ADJECTIVES, and finally feeds the clean tweets to the SVM and Naïve Bayes algorithm. In each approach, the SVM algorithm is shown to be more effective in identifying suspicious tweets than the Naïve Bayes algorithm.

4) Suspicious Twitter Posts Classification Implementation using SVM and Naive Bayes This module deals with inputting with SVM and Naive Bayes clean tweets and how the application will divide all the data into train and test datasets that will contain 80% and 20% of the data, respectively, first of all. At the initial stages, training of the algorithms will be done using 80% of the data and an appropriate model of the algorithm will be made. The model designed will be tested on test data to measure the accuracy, precision, recall, and F-Score of the predictions made.

5) Shut Up! SVM and CNN classification for gender and age prediction With S Kumar computing the trolling content then begins the tedious scanning process using the suspected troll tweet, and all images in a website if again above the waist and face segments will be obtained using an SVM classifier while the images put through the deep learning CNN will be to determine AGE and GEND.

6) Comparison Graph: in this module, we show a comparison graph between SVM and Naïve Bayes based on precision, recall and FSCORE**Output**

To execute the project, simply locate the 'run.bat' file and do a double click which will give rise to the screen below



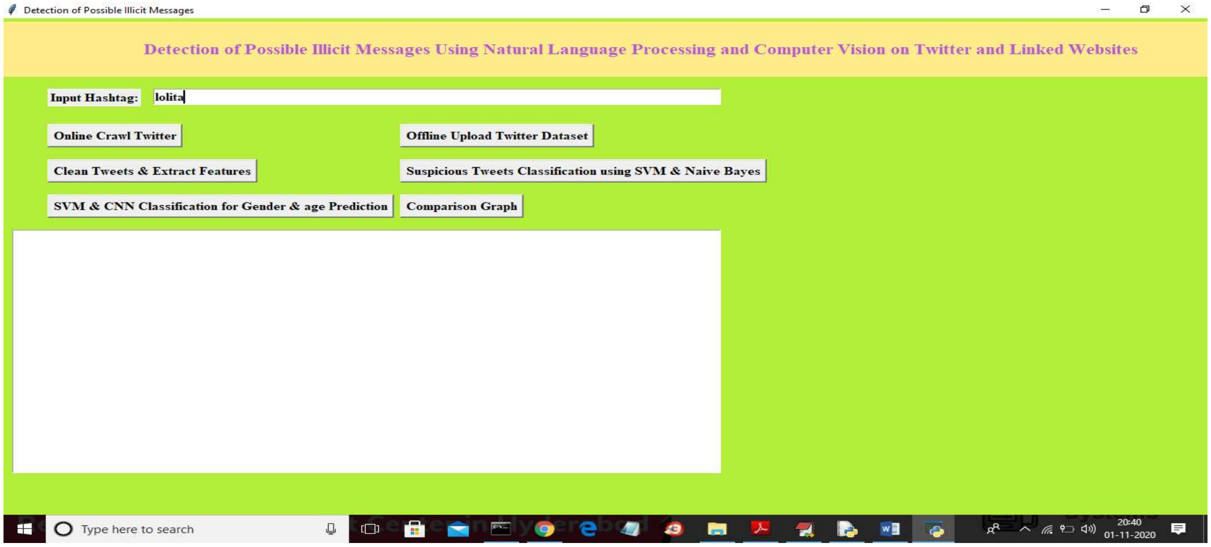
Detection of Possible Illicit Messages

Detection of Possible Illicit Messages Using Natural Language Processing and Computer Vision on Twitter and Linked Websites

Input Hashtag:

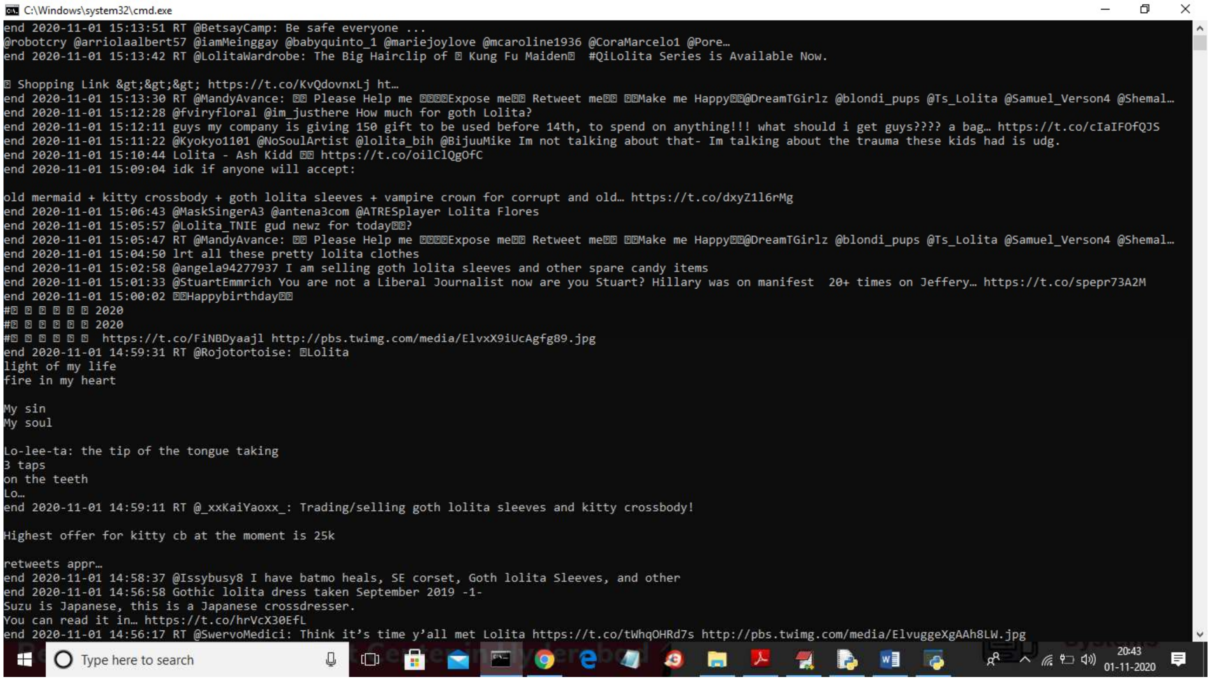
Online Crawl Twitter
Offline Upload Twitter Dataset
Clean Tweets & Extract Features
Suspicious Tweets Classification using SVM & Naive Bayes
SVM & CNN Classification for Gender & age Prediction
Comparison Graph
Type here to search
20:40
01-11-2020

On the above screen, please key in a hashtag and then press the ‘Online Crawl Twitter’ button to begin the process of crawling.

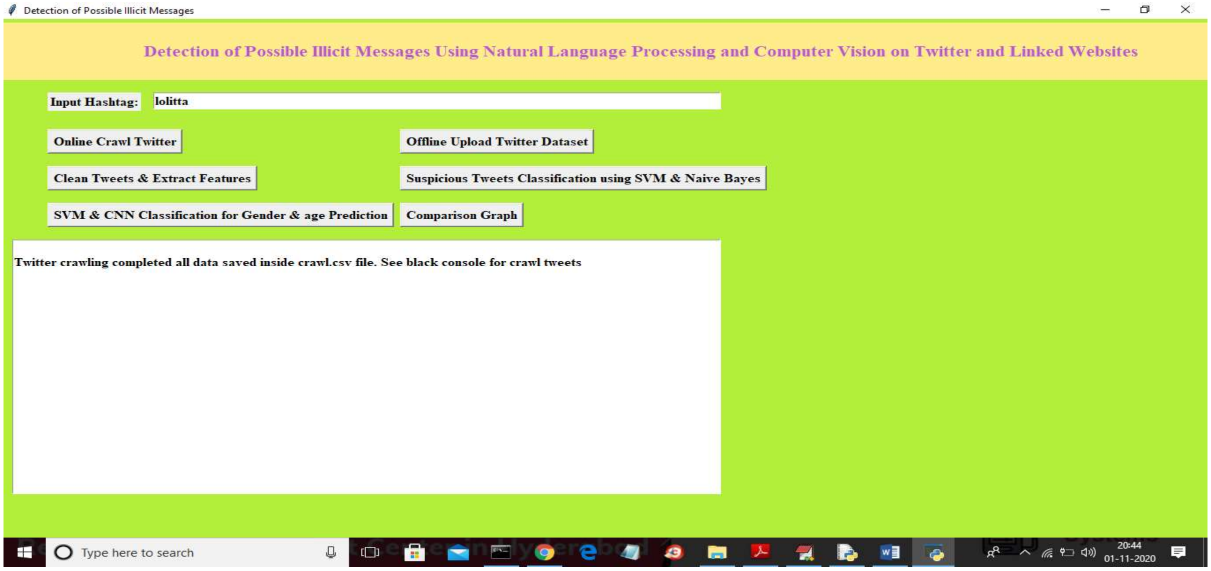

Detection of Possible Illicit Messages
Detection of Possible Illicit Messages Using Natural Language Processing and Computer Vision on Twitter and Linked Websites
Online Crawl Twitter
Offline Upload Twitter Dataset
Clean Tweets & Extract Features
Suspicious Tweets Classification using SVM & Naive Bayes
SVM & CNN Classification for Gender & age Prediction
Comparison Graph
Type here to search
20:40
01-11-2020

Input Hashtag: lolita

On the above screen, I input the hashtag ‘lolita’ and clicked the button ‘Online Crawl Twitter’ to begin the process of crawling. As can be observed in the black screen below, the crawling is in progress and I am displaying the date of the tweets and tweet text.



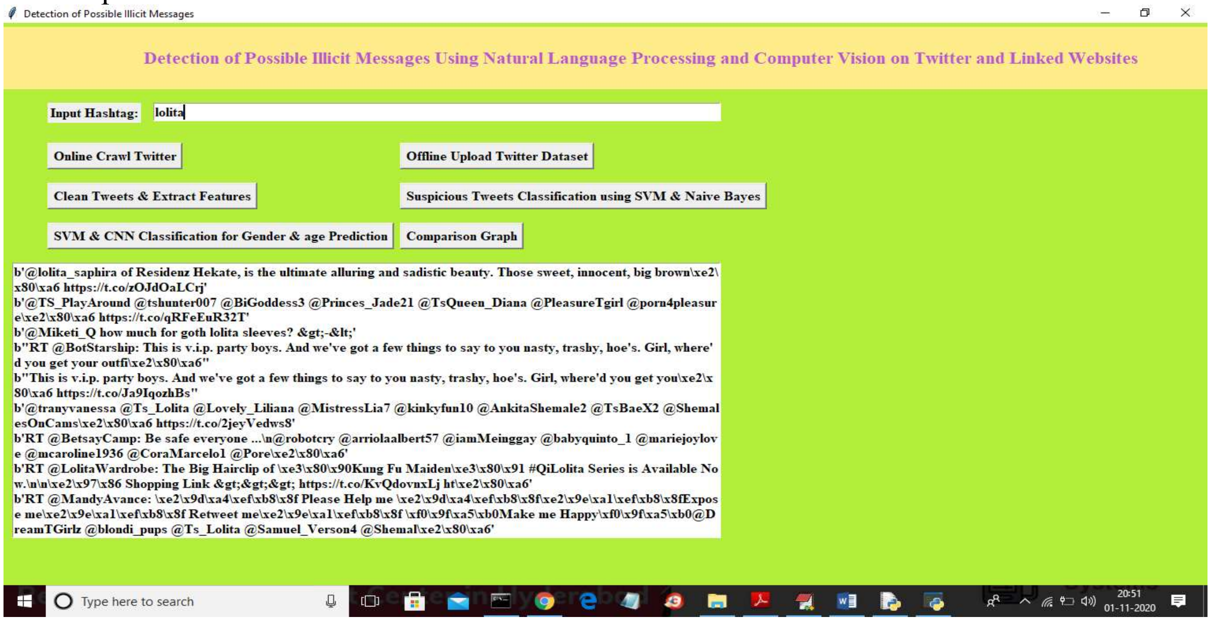
```
end 2020-11-01 15:13:51 RT @BetsayCamp: Be safe everyone ...
@robotcry @arriolaalbert57 @iamMeinggay @babyquinto_1 @mariejoylove @mcaroline1936 @CoraMarcelo1 @Pore…
end 2020-11-01 15:13:42 RT @LolitaWardrobe: The Big Hairclip of ▯ Kung Fu Maiden▯  #QiLolita Series is Available Now.

▯ Shopping Link &gt;&gt;&gt; https://t.co/KvQdovnxLj ht…
end 2020-11-01 15:13:30 RT @MandyAvance: ▯▯ Please Help me ▯▯▯▯Expose me▯▯ Retweet me▯▯ ▯▯Make me Happy▯▯@DreamTGirlz @blondi_pups @Ts_Lolita @Samuel_Verson4 @Shemal…
end 2020-11-01 15:12:28 @fviryfloral @im_justhere How much for goth Lolita?
end 2020-11-01 15:12:11 guys my company is giving 150 gift to be used before 14th, to spend on anything!!! what should i get guys???? a bag… https://t.co/cIaIFOfQJS
end 2020-11-01 15:11:22 @Kyokyo1101 @NoSoulArtist @lolita_bih @BijuuMike Im not talking about that- Im talking about the trauma these kids had is udg.
end 2020-11-01 15:10:44 Lolita - Ash Kidd ▯▯ https://t.co/oilClQgOfC
end 2020-11-01 15:09:04 idk if anyone will accept:

old mermaid + kitty crossbody + goth lolita sleeves + vampire crown for corrupt and old… https://t.co/dxyZ1l6rMg
end 2020-11-01 15:06:43 @MaskSingerA3 @antena3com @ATRESplayer Lolita Flores
end 2020-11-01 15:05:57 @Lolita_TNIE gud newz for today▯▯?
end 2020-11-01 15:05:47 RT @MandyAvance: ▯▯ Please Help me ▯▯▯▯Expose me▯▯ Retweet me▯▯ ▯▯Make me Happy▯▯@DreamTGirlz @blondi_pups @Ts_Lolita @Samuel_Verson4 @Shemal…
end 2020-11-01 15:04:50 lrt all these pretty lolita clothes
end 2020-11-01 15:02:58 @angela94277937 I am selling goth lolita sleeves and other spare candy items
end 2020-11-01 15:01:33 @StuartEmmrich You are not a Liberal Journalist now are you Stuart? Hillary was on manifest  20+ times on Jeffery… https://t.co/spepr73A2M
end 2020-11-01 15:00:02 ▯▯Happybirthday▯▯
#▯ ▯ ▯ ▯ ▯ ▯ ▯ 2020
#▯ ▯ ▯ ▯ ▯ ▯ ▯ 2020
#▯ ▯ ▯ ▯ ▯ ▯ ▯  https://t.co/FiNBDyaajl http://pbs.twimg.com/media/ElvxX9iUcAgfg89.jpg
end 2020-11-01 14:59:31 RT @Rojotortoise: ▯Lolita
light of my life
fire in my heart

My sin
My soul

Lo-lee-ta: the tip of the tongue taking
3 taps
on the teeth
Lo…
end 2020-11-01 14:59:11 RT @_xxKaiYaoxx_: Trading/selling goth lolita sleeves and kitty crossbody!

Highest offer for kitty cb at the moment is 25k

retweets appr…
end 2020-11-01 14:58:37 @Issybusy8 I have batmo heals, SE corset, Goth lolita Sleeves, and other
end 2020-11-01 14:56:58 Gothic lolita dress taken September 2019 -1-
Suzu is Japanese, this is a Japanese crossdresser.
You can read it in… https://t.co/hrVcX30EfL
end 2020-11-01 14:56:17 RT @SwervoMedici: Think it’s time y’all met Lolita https://t.co/tWhqOHRd7s http://pbs.twimg.com/media/ElvuggeXgAAh8LW.jpg
```


Detection of Possible Illicit Messages
Detection of Possible Illicit Messages Using Natural Language Processing and Computer Vision on Twitter and Linked Websites
Online Crawl Twitter
Offline Upload Twitter Dataset
Clean Tweets & Extract Features
Suspicious Tweets Classification using SVM & Naive Bayes
SVM & CNN Classification for Gender & age Prediction
Comparison Graph
Type here to search
20:44
01-11-2020

Input Hashtag: lolitta

```
Twitter crawling completed all data saved inside crawl.csv file. See black console for crawl tweets
```

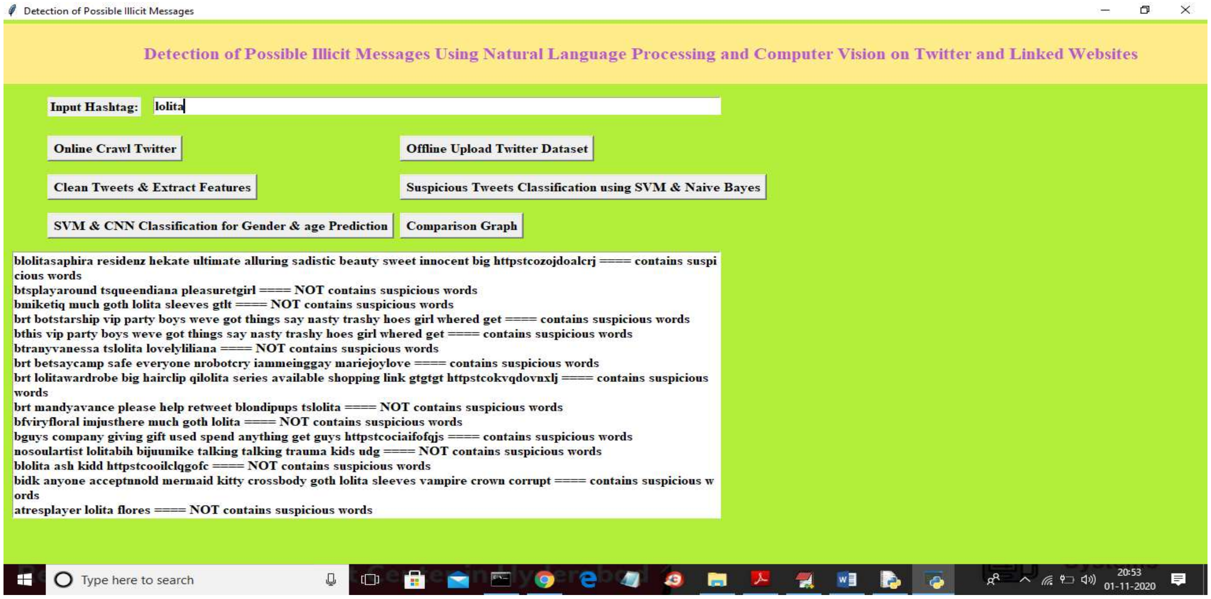
On the above screen, it can be observed that it would have shown ‘crawling completed’ over the Twitter platform and hence click on ‘Clean Tweets & Extract Features’ button now.


Detection of Possible Illicit Messages
Detection of Possible Illicit Messages Using Natural Language Processing and Computer Vision on Twitter and Linked Websites
Online Crawl Twitter
Offline Upload Twitter Dataset
Clean Tweets & Extract Features
Suspicious Tweets Classification using SVM & Naive Bayes
SVM & CNN Classification for Gender & age Prediction
Comparison Graph
Type here to search
20:51
01-11-2020

Input Hashtag: lolita

```
b'@lolita_saphira of Residenz Hekate, is the ultimate alluring and sadistic beauty. Those sweet, innocent, big brown\xe2\
x80\xa6 https://t.co/zOJdOaLCrj'
b'@TS_PlayAround @tshunter007 @BiGoddess3 @Princes_Jade21 @TsQueen_Diana @PleasureTgirl @porn4pleasur
e\xe2\x80\xa6 https://t.co/qRFeEuR32T'
b'@Miketi_Q how much for goth lolita sleeves? &gt;-&lt;'
b"RT @BotStarship: This is v.i.p. party boys. And we've got a few things to say to you nasty, trashy, hoe's. Girl, where'
d you get your outfi\xe2\x80\xa6"
b"This is v.i.p. party boys. And we've got a few things to say to you nasty, trashy, hoe's. Girl, where'd you get you\xe2\x
80\xa6 https://t.co/Ja9IqozhBs"
b'@tranyvanessa @Ts_Lolita @Lovely_Liliana @MistressLia7 @kinkyfun10 @AnkitaShemale2 @TsBaeX2 @Shemal
esOnCams\xe2\x80\xa6 https://t.co/2jeyVedws8'
b'RT @BetsayCamp: Be safe everyone ...\n@robotcry @arriolaalbert57 @iamMeinggay @babyquinto_1 @mariejoylov
e @mcaroline1936 @CoraMarcelo1 @Pore\xe2\x80\xa6'
b'RT @LolitaWardrobe: The Big Hairclip of \xe3\x80\x90Kung Fu Maiden\xe3\x80\x91 #QiLolita Series is Available No
w.\n\n\xe2\x97\x86 Shopping Link &gt;&gt;&gt; https://t.co/KvQdovnxLj ht\xe2\x80\xa6'
b'RT @MandyAvance: \xe2\x9d\xa4\xef\xb8\x8f Please Help me \xe2\x9d\xa4\xef\xb8\x8f\xe2\x9e\xa1\xef\xb8\x8fExpos
e me\xe2\x9e\xa1\xef\xb8\x8f Retweet me\xe2\x9e\xa1\xef\xb8\x8f \xf0\x9f\xa5\xb0Make me Happy\xf0\x9f\xa5\xb0@D
reamTGirlz @blondi_pups @Ts_Lolita @Samuel_Verson4 @Shemal\xe2\x80\xa6'
```

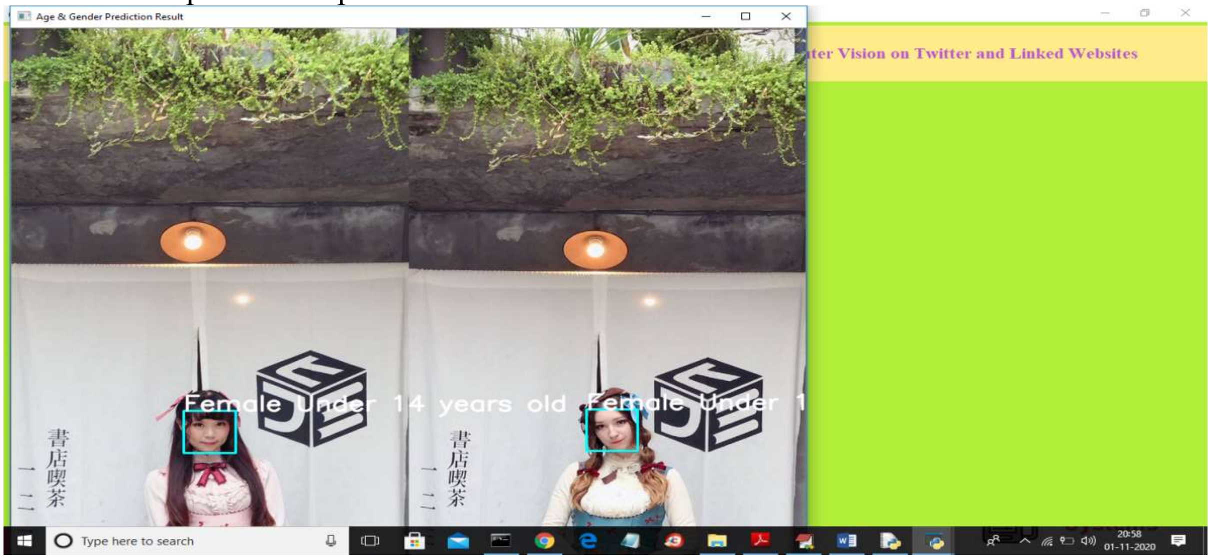
The above screen is showing all the clean tweets and display the result which is raw tweets before any cleaning process so that all the raw tweets processed for cleaning are shown and clean tweets are ready. Also, all the articles that have suspicious words click on ‘Suspicious Tweets Classification using SVM & Naive Bayes’ to apply SVM and Naïve Bayes on each of the tweet.


Detection of Possible Illicit Messages
Detection of Possible Illicit Messages Using Natural Language Processing and Computer Vision on Twitter and Linked Websites
Online Crawl Twitter
Offline Upload Twitter Dataset
Clean Tweets & Extract Features
Suspicious Tweets Classification using SVM & Naive Bayes
SVM & CNN Classification for Gender & age Prediction
Comparison Graph
Type here to search
20:53
01-11-2020

Input Hashtag: lolita

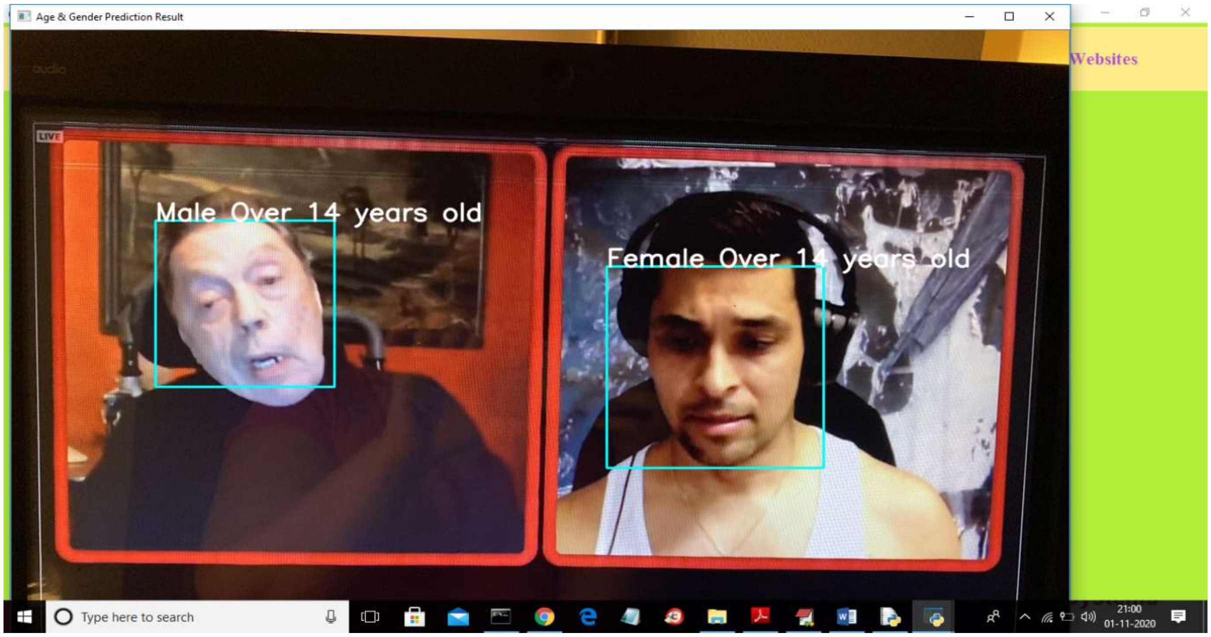
```
blolitasaphira residenz hekate ultimate alluring sadistic beauty sweet innocent big httpstcozojdoalcrj ==== contains suspi
cious words
btsplayaround tsqueendiana pleasuretgirl ==== NOT contains suspicious words
bmiketiq much goth lolita sleeves gtlt ==== NOT contains suspicious words
brt botstarship vip party boys weve got things say nasty trashy hoes girl whered get ==== contains suspicious words
bthis vip party boys weve got things say nasty trashy hoes girl whered get ==== contains suspicious words
btranyvanessa tslolita lovelyliliana ==== NOT contains suspicious words
brt betsaycamp safe everyone nrobotcry iammeinggay mariejoylove ==== contains suspicious words
brt lolitawardrobe big hairclip qilolita series available shopping link gtgtgt httpstcokvqdovnxlj ==== contains suspicious
words
brt mandyavance please help retweet blondipups tslolita ==== NOT contains suspicious words
bfviryfloral imjusthere much goth lolita ==== NOT contains suspicious words
bguys company giving gift used spend anything get guys httpstcociaifofqjs ==== contains suspicious words
nosoulartist lolitabih bijuumike talking talking trauma kids udg ==== NOT contains suspicious words
blolita ash kidd httpstcooilclqgofc ==== NOT contains suspicious words
bidk anyone acceptnnold mermaid kitty crossbody goth lolita sleeves vampire crown corrupt ==== contains suspicious w
ords
atresplayer lolita flores ==== NOT contains suspicious words
```

In above screen showing each cleaned tweets and after equals to sign showing detected result which contains suspicious words or not. In this way we got tweets which has suspicious words and now click on ‘SVM & CNN Classification for Gender & age Prediction’ button to go through each tweets website to see picture and predict AGE and GENDER.

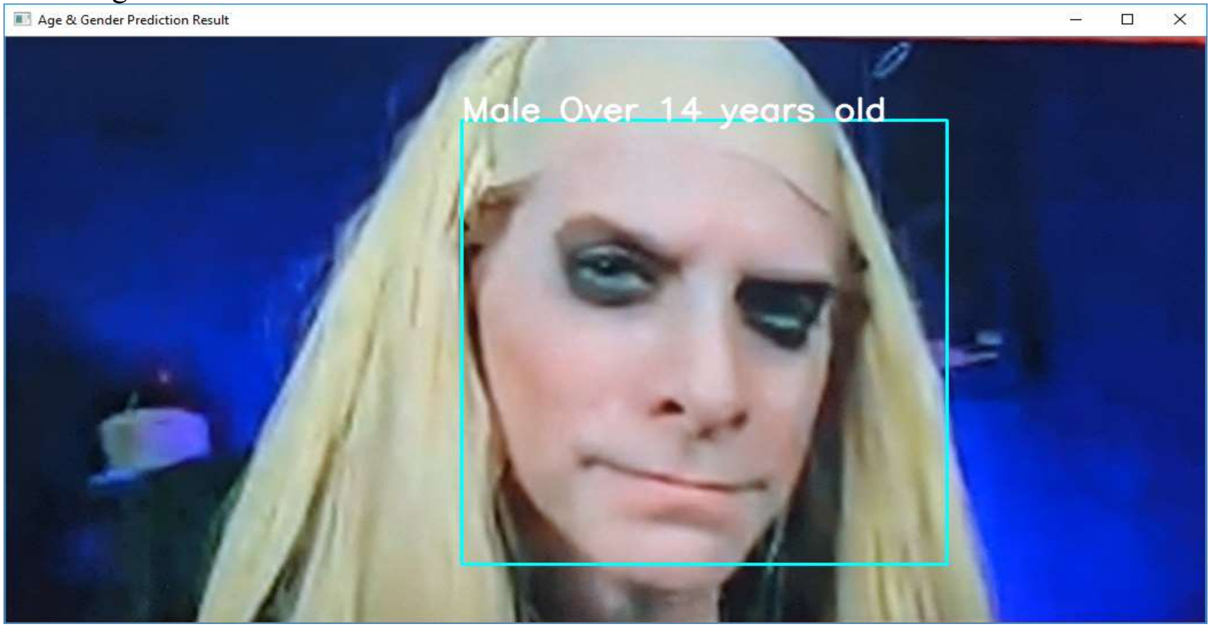

Age & Gender Prediction Result
ter Vision on Twitter and Linked Websites
Female Under 14 years old
Female Under 1
書店喫茶
書店喫茶
Type here to search
20:58
01-11-2020

For this figure, within the application, a detected image of a face has been portrayed as being female and under the age of 14 years. The application performs the same process for all the tweets. …turn to the next figure below.

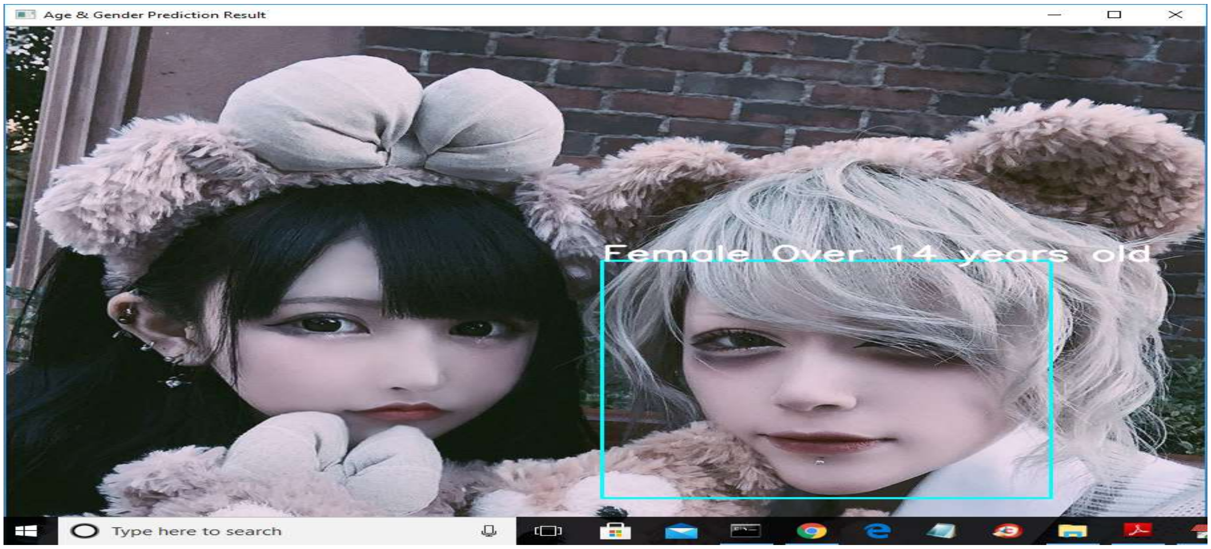
ISSN:1624-1940
DOI 10.6084/m9.figshare.2632599
http://magellanes.com/



Above figure is another result of this screenshot.

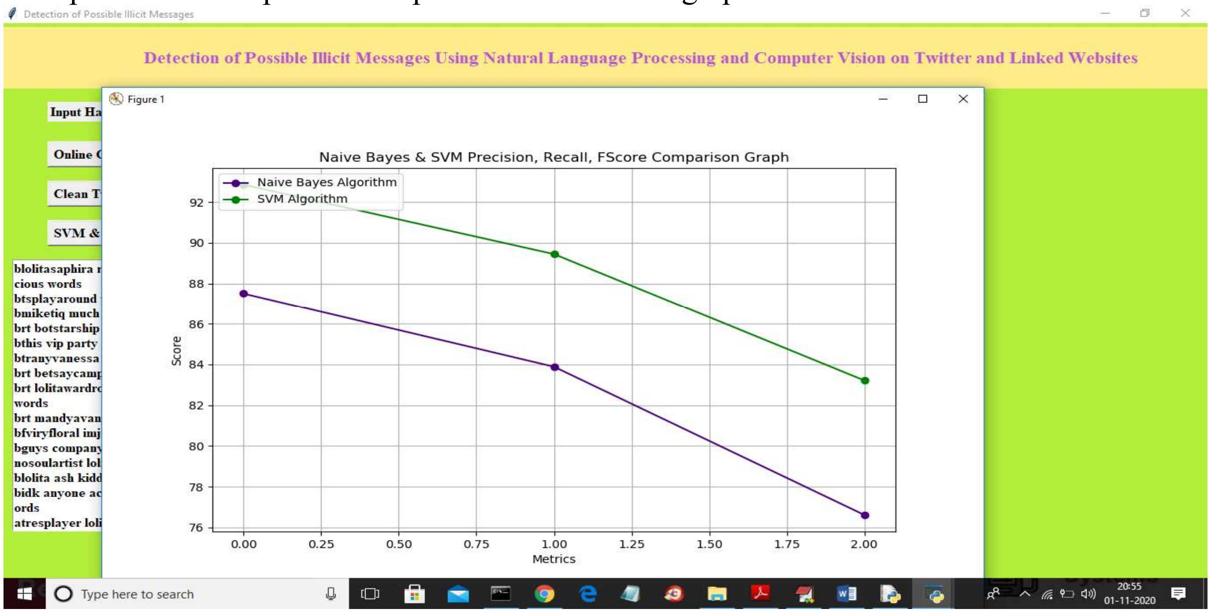


ISSN:1624-1940
DOI 10.6084/m9.figshare.2632599
http://magellanes.com/



Note: No algorithm is ever truly capable of 100% detection of accurate face and gender in any image, thus our algorithms also will provide a 70% correct prediction as I have not trained the algorithm neural network with a large number of epochs. We cannot accomplish the training of a system of such database with an outlay of time and resources of just a few hundred epochs. This is why the model will only give 70% accurate predictions.

Now press the 'Comparison Graph' button to see the graph below.



blolitasaphira r
cious words
btsplayaround
bmiketiq much
brt botstarship
bthis vip party
btranyvanessa
brt betsaycamp
brt lolitawardrc
words
brt mandyavan
bfviryfloral imj
bguys company
nosoulartist lol
blolita ash kidd
bidk anyone ac
ords
atresplayer loli

In the presented graph, the blue color indicates the values of precision, recall and FScore based on Naïve Bayes while the green one illustrates those based on SVM. In the x-axis of the presentation above, the components precision, recall and F-score are represented while their respective values are shown in the y-axis. Therefore from the above graph it can be suggested that SVM is performing the best.**Conclusion**

Face recognition algorithms and machine learning models have undergone some improvements over the past couple of years. As such, in the ILSVRC competition level 90% accuracy ± 5% was achieved. To such an extent, one can claim that the machine learning recognition can be likened to images recognition performed at ease by a human. Image recognition is influenced directly by various elements such as size, color, opacity, resolution, type of image format, among others. Thus image recognition and classification performance depends largely on the quality of dataset used.

In this research, we studied that, encouraging results can be achieved employing only the geometric features of the trunk without using a solely facial features. In this subset, Haar-like features were extracted and employed into an SVM classifier, then age and gender were classified using an SVM. Found results were obtained along with those of a CNN algorithm.

SVM is one accepted model, and in this study, we achieved a classification rate of more than 80% for two experiments (face as well as upper body) not only in gender classification but also in age group classification in this work. The significant contribution in this paper is based on upper body image classification to detect age group in order to combat human trafficking.o the best of understanding, this work is the first attempt of its kind regarding image classification with no face features, instead using just their upper body shape. No such work exists right now

Future Work

To conclude, future works contain: 1) analysis of certain aspects of ethnic and race characteristics, 2) to broaden the initiative in order to derive geometric aspects of full body as well as different types of images or comprehensive videos in diverse formats, 3) identification of health risks through processing of images that contain various features of torso, legs, and back among others, and 4) implementation of other algorithms in the context of other networks for example Instagram..

References:

[1] B. Bangerter, S. Talwar, R. Arefi, and K. Stewart, ‘‘Networks and devices for the 5G era,’’ IEEE Commun. Mag., vol. 52, no. 2, pp. 90–96, Feb. 2014.

[2] F. Laczko, ‘‘Data and research on human trafficking,’’ Int. Migration, vol. 43, nos. 1–2, pp. 5–16, Jan. 2005.

[3] M. Lee, ‘‘Human trafficking and border control in the global south,’’ in The Borders of Punishment: Migration, Citizenship, and Social Exclusion. Oxford, U.K.: Oxford Univ. Press, 2013, pp. 128–149.

[4] E. Cockbain and E. R. Kleemans, ‘‘Innovations in empirical research into human trafficking: Introduction to the special edition,’’ Crime, Law Social Change, vol. 72, no. 1, pp. 1–7, Jul. 2019.

[5] R. Weitzer, ‘‘Human trafficking and contemporary slavery,’’ Annu. Rev. Sociol., vol. 41, pp. 223–242, Aug. 2015.

[6] T. S. Portal. (2018). Twitter: Number of Monthly Active Users 2010-2018. [Online]. Available: https://www.statista.com

[7] M. R. Candes, ‘‘The victims of trafficking and violence protection act of 2000: Will it become the thirteenth amendment of the twenty-first century,’’ U. Miami Inter-Amer. L. Rev., vol. 32, p. 571, Jun. 2001.

[8] D. Hughes, Wilberforce Can be Free Again: Protecting Trafficking Victims. New York, NY, USA: National Review, 2008.

[9] A. Sultan, ‘‘Countering crime trafcking in persons smuggling migrants Ethiopia: The Law practice,’’ Ph.D. dissertation, School Law, Addis Ababa Univ., Ababa, Ethiopia, 2018, pp. 1–72.

[10] M. Tsikerdekis and S. Zeadally, ‘‘Online deception in social media,’’ Commun. ACM, vol. 57, no. 9, pp. 72–80, Sep. 2014.

[11] A. Vishwanath, ‘‘Diffusion of deception in social media: Social contagion effects and its antecedents,’’ Inf. Syst. Frontiers, vol. 17, no. 6, pp. 1353–1367, Jun. 2014.

[12] E. Tong, A. Zadeh, C. Jones, and L.-P. Morency, ‘‘Combating human trafficking with deep multimodal models,’’ 2017, arXiv:1705.02735. [Online]. Available: http://arxiv.org/abs/1705.02735

[13] J. V. D. Wolfshaar, M. F. Karaaba, and M. A. Wiering, ‘‘Deep convolutional neural networks and support vector machines for gender recognition,’’ in Proc. IEEE Symp. Ser. Comput. Intell., Dec. 2015,

ISSN:1624-1940
DOI 10.6084/m9.figshare.2632599
http://magellanes.com/

pp. 188–195.
[14] M. Hernandez-Alvarez,‘‘Detection of possible human trafficking in Twitter,’’ in Proc. Int. Conf. Inf. Syst. Softw. Technol. (ICIST), Nov. 2019, pp. 187–191.
[15] H. Alvari, P. Shakarian, and J. E. K. Snyder, ‘‘A non-parametric learning approach to identify online human trafficking,’’ in Proc. IEEE Conf. Intell. Secur. Informat. (ISI), Sep. 2016, pp. 133–138.
[16] M. M. Dehshibi and A. Bastanfard, ‘‘A new algorithm for age recognition from facial images,’’ Signal Process., vol. 90, no. 8, pp. 2431–2444, Aug. 2010.
[17] F. Salvetti, ‘‘Detecting deception in text: A corpus-driven approach,’’ Ph.D. dissertation, Comput. Sci. Graduate, Univ. Colorado Boulder, Boulder, CO, USA, 2012, pp. 1–206.
[18] S. Sarkar, ‘‘Use of technology in human trafficking networks and sexual exploitation: A cross-sectional multi-country study,’’ Trans. Social Rev., vol. 5, no. 1, pp. 55–68, Jan. 2015.