Detection of Potentially Illicit Messages

A Project Work Synopsis

Submitted in the partial fulfilment for the award of the degree of

BACHELOR OF ENGINEERING

IN BIG DATA AND ANALYTICS

Submitted by:

Vaibhav Kumar Singh
20BCS3842

Under the Supervision of:

Dr. Alankrita Aggarwal

CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413, PUNJAB

07 MARCH, 2023

ABSTRACT

Millions of victims of human trafficking suffer from this widespread issue, which robs them of their dignity. Social networks are currently utilised to disseminate this crime across the internet by disseminating hidden messages that advertise these illicit services. It is crucial in this situation to automatically discover messages that may be connected to this crime and could possibly act as clues because law enforcement resources are limited. In this study, we use natural language processing to find tweets that might advertise these illicit services and take advantage of children. It is feasible to find pictures of children under 14 years old since the images and URLs obtained in suspicious messages were processed and categorised by gender and age**.**The approach we took was as follows. First, real-time hashtag mining is done on tweets referencing minors. These tweets are preprocessed to remove background noise and typos before being categorised as suspicious or not. Moreover, Haar models are used to choose the geometrical characteristics of the face and body. We can identify a person's gender and age group by using support vector machines (SVM) and convolutional neural networks (CNN), even when the details of the face are blurry or the torso is out of proportion to the head. As a result, the SVM model with torso-only features performs better than CNN.

Keywords- Support Vector Machine (SVM) and Convolutional Neural Network (CNN).

CHANDIGARH UNIVERSITY
Discover. Learn. Empower.

Table of Contents

List of Tables

1.INTRODUCTION

As the user could not actually interact with the web at first, the websites were segregated and just intended for reading. Yet with the invention and introduction of web 2.0, there was a revolutionary and radical transformation since users ceased to be passive bystanders and started participating actively in social networks like Facebook, Twitter, and Instagram, among others.

A door has unfortunately also been opened for criminal operations like human trafficking, where some nations, including those in South America, have the greatest rates of smuggling of persons, particularly children and teenagers under the age of 14. In recent years, numerous criminal groups have advertised these "sexual services" on social media while concealing their unlawful behaviour with words like "chicken soup" to refer to child pornography. Websites and social media platforms are utilised to bring this crime into the online world, where covert messaging and advertising are used to advertise illicit services and take advantage of the victims of this crime, who are primarily kids. Although there have been prior attempts to identify unlawful messages using picture classification and tweet filtering, most of these systems combine computer vision and natural language processing. It is demonstrated that text and graphics are handled differently. The writers of this research concentrate their efforts on the analysis of online advertisements for automatic message detection.

PROBLEM DEFINITION

The main focus of this project is to use training data, specifically the data from law enforcement, to create a Human Illicit message detection system. One of the most difficult issues for law enforcement to combat is human trafficking, which requires global cooperation. In this work, we use easily accessible data from the classified ads website "Backpage" to identify the most likely adverts related to human trafficking and uncover potential patterns of online human trafficking operations.

1.1 PROJECT OVERVIEW

- Using predictive models, such as Vector Support Machine (SVM) and Convolutional Neural Networks (CNN), the image classification process is done through a training phase and a testing phase.The data for the Sales will be taken from Kaggle.
- A text-based deception detection model for cybercrime

- Prediction will be automatedwith Machine learning Models thus saving a lot of time.
- We will be working on Classification, Pandas, Matplotlib, Model Training, SVM and Naïve Bayes machine learning algorithms.

1.2 HARDWARE SPECIFICATIONS

1.3.1 PC

A pc is a personal computer that can be used for multiple purposes depending on its size, capabilities, and price. They are to be operated directly by the end-user. Personal computers are single-user systems and are portable. Our web application program will be installed on the pc for our clients to use it. This makes it feasible for individual use.

1.3 SOFTWARE SPECIFICATIONS

1.4.1 Jupyter Notebook:

Jupyter Notebook is a web-based open-source application that is used for editing, creating running, and sharing documents that contain live codes, visualization, text, and equations. Its core supported programming languages are Julia, R, and Python. Jupyter notebook comes with an IPython kernel that allows the programmer to write programs in python. There are over 100 kernels other than IPython available for use.

1.4.2 Atom Text editor

Atom is a text and source code editor which works across all operating systems. It speeds up find-and-replace operations by an order of magnitude and improves loading performance for large, single-line files It's a desktop application built with HTML, JavaScript, CSS, and Node.js integration.

1.4.3 SVM

A supervised machine learning approach called Support Vector Machine (SVM) is used for both classification and regression. Although we also refer to regression concerns, categorization is the most appropriate term. Finding a hyperplane in an N-dimensional space that clearly classifies the data points is the goal of the SVM method. The number of features

determines the hyperplane's size.

The hyperplane is essentially a line if there are just two input features. The hyperplane turns into a 2-D plane if there are three input features. Imagining something with more than three features gets challenging. The SVM kernel is a function that converts non separable problems into separable problems by taking low-dimensional input space and transforming it into higher-dimensional space. It works best in non-linear separation issues. Simply explained, the kernel determines how to split the data depending on the labels or outputs defined after performing some incredibly sophisticated data transformations.

1.4.4 CNN

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning method that can take in an input image, give various elements and objects in the image importance (learnable weights and biases), and be able to distinguish between them. Because of its great accuracy, CNNs are utilised for picture categorization and recognition. Yann LeCun, a computer scientist, first proposed it in the late 1990s after becoming intrigued by how humans recognise objects visually. The principal applications of a convolutional neural network (CNN), which comprises one or more convolutional layers, are image processing, classification, segmentation, and other auto correlated data.

2.LITERATURE REVIEW

One of the most difficult issues for law enforcement to combat is human trafficking, which requires global cooperation. In this work, we use easily accessible data from the classified ad website "Backpage" to determine the most likely ads that are related to human trafficking and uncover potential patterns of online human trafficking operations. We rely on two human analysts—one human trafficking victim survivor and one from law enforcement—to manually classify the little amount of crawled data because there isn't any ground truth. Then, we introduce a semi-supervised learning strategy that is trained on the available labelled and unlabeled data and evaluated on unobserved data with additional expert verification.

2.1 Existing System Summary

Although there are previous tweet filtering and image classification works to detect illicit messages, most of them use natural language processing methods or computer vision techniques. Semi-structured questionnaires were used in the survey to collect information about how victims used technology before and after being trafficked, how they advertised themselves, the various services and technologies that were used to trade in sexually exploited trafficked people, and how clients looked for, communicated with, and paid for their sex experiences. The findings demonstrated that traffickers and

their networks effectively utilised sophisticated software to protect their anonymity, made use of internet storage and hosting services, and employed cutting-edge encryption methods to thwart police digital forensic investigations. We suggest that these hopeful results in automatic deception detection are primarily the consequence of the side effects of corpus-specific features because our aim is in discovering the invariant properties of deceit in text. Although it doesn't hinder practical uses, this does not encourage a more thorough study of dishonesty. We have created the largest publicly accessible shared multidimensional deception corpus for online reviews, the BLT-C (Boulder Lies and Truths Corpus), to illustrate this and enable researchers and practitioners to share results. We have also created a collection of semi-automatic algorithms to guarantee corpus validity in an effort to get over the inherent lack of ground truth.This thesis demonstrates how difficult it is to identify deceit using supervised machine learning techniques. Millions of victims of human trafficking suffer from this widespread issue, which robs them of their dignity. Social networks are currently utilised to disseminate this crime across the internet by disseminating hidden messages that advertise these unlawful services. It is crucial in this situation to automatically discover messages that may be connected to this crime and could possibly act as clues because law enforcement resources are limited. The prevalence of text messages and cybercrime using text-based deceptive discourse are both rising. In cybercriminal networks, we utilise machine learning and linguistic techniques to identify dishonesty in text communications. Through web genre, we create cybercrime detection models. Our contributions include the following: models trained on fraud in the email genre can anticipate frauds in the social media web genre with 50% predictive accuracy; models trained on fraud in the social media web genre can detect fraud in communications in the email web genre with 60% predictive accuracy. Due to the linguistic diversity of the hackers in this study, the prognosis for the email model is optimistic. We also show that elements from natural language processing and linguistic psychological processes related to cybercrime can be used to build cybercrime detection models.

Year and citation	https://www.researchgate.net/publication/339547729_Detection_of_Possible_Illicit_Messages_Using_Natural_Language_Processing_and_Computer_Vision_on_Twitter_and_Linked_Websites march 2021.	Niloufar Shoeibi, Nastaran Shoeibi, Guillermo Hernández, Pablo Chamoso, J. Corchado less Published 1 November 2022	https://www.academia.edu/58265879/Detection_of_Possible_Illicit_Messages_Using_Natural_Language_Processing_and_Computer_Vision_on_Twitter_and_Linked_Websites?from_sitemaps=true&version=2 Published 2023
Article Title	Detection of Possible Illicit Messages Using Natural Language Processing.	AI-Crime Hunter: An AI Mixture of Experts for Crime Discovery on Twitter	Computer Vision on Twitter and Linked Websites
Purpose of the study	Human trafficking is a global problem that strips away the dignity of millions of victims. Currently, social networks are used to spread this crime through the online environment by using covert messages that serve to promote these illegal services.	Maintaining a healthy cyber society is a big challenge due to the users' freedom of expression and behaving. It can be solved by monitoring and analyzing the users' behavior and taking proper actions towards them. This research aims to present a platform that monitors the public content on Twitter by extracting tweet data.	The images and the URLs found in suspicious messages were processed and classified by gender and age group, so it is possible to detect photographs of people under 14 years of age. The method that we used is as follows. First, tweets with hashtags related to minors are mined in real-time. These tweets are preprocessed to eliminate noise and misspelled words, and then the tweets are classified as suspicious or not.
Tools/ Software used	- Jupyter Notebook	- Jupyter Notebook	- Jupyter Notebook
Comparison of techniques done	- Decision Tree (DT) - Image Classification - Deep Belief Networks (DBNs)	- DistilBERT - FFNN	- SVM -CNN -Lasso
Evaluation parameters	- Model Accuracy	- Model Accuracy	- Model Accuracy

Table 2.2: Literature review summary

2.2 Proposed System

In this research, the author presents a concept to identify human trafficking through the analysis of social media text messages using SVM and Naive Bayes machine learning techniques. Using words like Lolita, escort, and others, the author of this paper first crawls Twitter. The extracted tweets are then cleaned to remove special symbols and stop words (words like the, where, and, an, etc.), and then they are analysed to extract words like VERBS and ADJECTIVE, which may contain important subjects or suspicious words used by HUMAN TRAFFICKERS (the suspicious words can include chicken soup, girls, penguin, and more).
SVM and Nave Bayes classifiers will be fed clean tweets as input to find suspicious terms.

3. PROBLEM FORMULATION

- Online Twitter Crawl: Using the TWEEPY API, this module allows us to enter a hashtag and have the program search Twitter for any tweets containing that keyword.
- Offline Upload Twitter Dataset: In this module, you can upload an existing Twitter dataset if you don't want to crawl Twitter.
- Clean Tweets & Extract Features: With this module, each tweet is processed to remove special symbols and stop words before being extracted of its verbs and adjectives, which are then fed into the SVM and Naive Bayes algorithms. SVM provides a superior result for the detection of suspicious tweets than the Naive Bayes algorithm.
- Suspicious Tweets Classification Using SVM & Naive Bayes: In this module, we input clean tweets to the SVM and Naive Bayes algorithms. The program then divides the complete data set into train and test sections, utilising 80% of the data for training and 20% of the data for testing. Algorithms will first be trained and a model will be generated using 80% of the data. Using test data, a trained model will be used to determine prediction accuracy, precision, recall, and FSCORE.
- SVM & CNN Classification for Gender & Age Prediction: Following the discovery of suspicious tweets, each suspicious tweet website is scanned to read all photos, and from those images, the face and upper body parts are extracted using SVM classifier.

4. OBJECTIVES

The proposed work is aimed to carry out work leading to the development of an approach for preventing Human Trafficking. The system, which is theproposed work will be achieved by dividing the work into the following objectives:

1. Obtain the Big Maty dataset from the web
2. Perform Data Wrangling.
3. Work on different Machine learning Algorithms
4. Try to develop algorithms based on different models to achieve maximum accuracy.
5. To develop the model using SVM and deploy the model using CNN, A Deep Learning Algorithm.

5. METHODOLOGY

The following methodology will be followed to achieve the objectives defined for the proposed research work:

1. Detailed study on Human Trafficking will be done.
2. Installation and hands-on experience on existing approaches of Human Trafficking will be done. Relative pros and cons will be identified.
3. Various parameters will be identified to evaluate the proposed system.
4. Comparison of newly implemented approach with existing approaches will be done.

6. TENTATIVE CHAPTER PLAN FOR THE PROPOSED WORK

CHAPTER 1: INTRODUCTION

This chapter will cover the overview of Human Trafficking and the different machinelearning and deep learning algorithms.

CHAPTER 2: LITERATURE REVIEW

This chapter includes the literature available for Human Trafficking The findings of the researchers will be highlighted which will become the basis of the current implementation.

CHAPTER 2: BACKGROUND OF PROPOSED METHOD

This chapter will provide an introduction to the concepts which are necessary to understand the proposed system.

CHAPTER 4: METHODOLOGY

This chapter will cover the technical details of the proposed approach.

CHAPTER 5: EXPERIMENTAL SETUP

This chapter will provide information about the subject system and tools used for the evaluation of the proposed method.

CHAPTER 6: RESULTS AND DISCUSSION

The result of the proposed technique will be discussed in this chapter.

CHAPTER 7: CONCLUSION AND FUTURE SCOPE

The major finding of the work will be presented in this chapter. Also, directions for extending the current study will be discussed.

7. 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.