CHANDIGARH UNIVERSITY
Discover. Learn. Empower.

DETECTION OF POSSIBLE ILLICIT MESSAGES USING NATURAL LANGUAGE PROCESSING AND COMPUTER VISION ON TWITTER

A Project Work Report

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

10 JANUARY, 2023

CHANDIGARH UNIVERSITY
Discover. Learn. Empower.

BONAFIDE CERTIFICATE

Certified that this project report "Detection of Possible Illicit Messages UsingNatural Language Processing and Computer Vision on Twitter" is the bonafide work of "Vaibhav Kumar Singh" who carried out the project work under my/our supervision.

SIGNATURE

SIGNATURE

HEAD OF THE DEPARTMENT

SUPERVISOR

Submitted for the project viva-voce examination held on _

INTERNAL EXAMINER

EXTERNAL EXAMINER

CHANDIGARH UNIVERSITY
Discover. Learn. Empower.

ACKNOWLEDGEMENT

I would like to thank my supervisor, Dr.Alankrita Aggarwal Mam, for her guidance and advice through each stage of making this project and giving me Vaibhav Kumar Singh, this opportunity to work on a minor project in which we can show our true potential, creativity, and hard work.

I would also like to thank my family and friends, who have been a constant support and have always motivated me to work hard and bring out the best in me.

This project’s success and end necessitated a great deal of direction and assistance from many people, and we are extremely fortunate to have received it all as part of the project's completion.

We owe everything we've accomplished to their oversight and help, and we'd like to express our gratitude.

ABSTRACT

This global problem, which robs millions of victims of human trafficking of their dignity, causes them to suffer. Currently, social networks are used to spread this criminal activity over the internet by broadcasting covert messages that promote these illegal businesses. Due of the limited resources available to law enforcement, it is critical in this case to automatically find texts that may be connected to this crime and may serve as clues. In this study, we search for tweets that potentially promote these illegal businesses and exploit children using natural language processing. Since the photographs and URLs found in suspicious communications were processed and categorised by gender and age, it is possible to find images of children under the age of 14. The method we used was as follows. First, tweets mentioning minors are mined for real-time hashtags. Before being classified as suspicious or not, these tweets are preprocessed to remove background noise and errors. Furthermore, the geometrical features of the body and face are selected using Haar models. Support vector machines (SVM) and convolutional neural networks (CNN) can be used to determine a person's gender and age group, even if the face's details are hazy or the torso is out of proportion to the head. Because of this, the SVM model with torso-only features outperforms CNN.

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

CHANDIGARH UNIVERSITY
Discover. Learn. Empower.

Table of Contents

List of Tables

1. INTRODUCTION

The websites were divided and only meant for reading at first because the user could not actively engage with the web. However, the creation and adoption of web 2.0 brought about a revolutionary and radical transition as users stopped acting as passive spectators and began actively engaging in social networks like Facebook, Twitter, and Instagram, among others.

Unfortunately, a door has also been created for illegal activities like human trafficking. Some countries, notably those in South America, have the highest rates of people smuggling, especially children and teens under the age of 14. Numerous criminal organisations have promoted these so called wrong services in recent years on social media while obscuring their illegal actions by using terms like "chicken soup" to refer to child pornography. Websites and social media platforms are utilised to spread this crime online, where covert messaging and advertising are used to promote illegal services and exploit the victims, who are mainly children. The majority of these methods involve computer vision and natural language processing, notwithstanding earlier attempts to identify illicit messages using picture categorization and tweet filtering. It is shown that the treatment of text and visuals differs. The authors of this study focus their efforts on analysing web adverts to find messages automatically.

PROBLEM DEFINITION

The major goal of this research is to develop a system for detecting human illicit messages using training data, specifically data from law enforcement. Human trafficking is one of the most challenging problems for law enforcement to address and calls for international cooperation. In this study, we identify the most likely advertisements related to human trafficking and identify potential patterns of online human trafficking operations using readily available data from the classified ads website "Backpage".
This global problem, which robs millions of victims of human trafficking of their dignity, causes them to suffer. Currently, social networks are used to spread this criminal activity over the internet by broadcasting covert messages that promote these illegal businesses.
The method we used was as follows. First, tweets mentioning minors are mined for real-time hashtags. Before being classified as suspicious or not, these tweets are preprocessed to remove background noise and errors.

1.1 PROJECT OVERVIEW

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

- Prediction Machine learning models will automate prediction, saving a tonne of time.

- We'll be working with the machine learning algorithms SVM, Pandas, Matplotlib, Model Training, and Nave Bayes.

The project overview shows that there are various phases involved in the detection of potentially illegal communications on Twitter utilising natural language processing (NLP) and computer vision techniques. The project overview is explained in more detail below:

Clarifying the problem we're attempting to tackle is the first step in any data science effort. The challenge in this situation is to use NLP and computer vision techniques to identify potentially hazardous or incorrect content on Twitter.

Data Gathering: The following stage is gathering a dataset of tweets that are known to include illegal messages. This can entail looking for tweets that have been reported as offensive or utilising terms that are frequently connected to illegal content. To make sure that our model is not biassed, we need also take into account any potential biases in the dataset and take action to address them.

Data preprocessing: Once the dataset has been gathered, the data needs to be prepared for analysis. Stop words, punctuation, and other distracting elements from the text must be eliminated, and the data must then be further refined using stemming, lemmatization, and tokenization methods. Additionally, we must use computer vision algorithms to preprocess any pictures or videos that are posted on Twitter.

Natural Language Processing (NLP): Using the preprocessed data as a starting point, NLP techniques can be used to recognise and classify tweets. Using named entity identification to locate significant entities like people, locations, and organisations, sentiment analysis to determine the emotional tone of the tweet, topic modelling to categorise tweets into categories, and sentiment analysis to determine. The data showed patterns.

Computer Vision: Using computer vision techniques, we can examine both the text and the photos and videos uploaded on Twitter. Using machine learning algorithms, this entails classifying pictures or movies as possibly illegal by spotting patterns in the visual data. For instance, we can utilise object recognition to find pictures of criminal activity, guns, or drugs.

Feature Extraction: After preprocessing the data and utilising NLP and computer vision techniques to extract features, we can utilise these features to train a machine learning model to categorise tweets as possibly unlawful or not. This could entail identifying trends in the data and making assumptions about whether a specific tweet is appropriate or not using methods like logistic regression, decision trees, or neural networks.

Model Training: Using the preprocessed data and the machine learning algorithm we choose after extracting the features. In this case, the data would be divided into training and testing sets, with the training set being used to train the model and the testing set to assess how well it performed.

Model Evaluation: After our machine learning model has been trained, we must assess its effectiveness in identifying potentially illegal texts. This could entail evaluating the model's accuracy using measures like precision, recall, and F1 score as well as performing a human evaluation of the flagged tweets to determine whether or not they are actually unsuitable.

Finally, we must put our system into operation and integrate it with a monitoring tool that can continuously search Twitter for potentially hazardous content. Building a web application that enables human moderators to examine flagged tweets and take appropriate action, like deleting the tweet or banning the user who tweeted it, is one way to accomplish this.

A machine learning model is trained to categorise tweets as potentially illicit or not, its performance is assessed, and the system is then deployed for continuous monitoring of Twitter. The project overview for detecting possible illicit messages on Twitter using NLP and computer vision techniques includes defining the problem, collecting and preprocessing the relevant data, applying NLP and computer vision techniques to extract features from the data, and training the model. This initiative has the ability to increase the security and safety of Twitter users by quickly and effectively identifying and eliminating offensive or inappropriate content.

1.2 HARDWARE SPECIFICATIONS

1.3.1 PC

A personal computer, or pc, is a device that, depending on its dimensions, features, and cost, may be used for a variety of tasks. The end user is to operate them directly. Single-user, portable systems are personal computers. For our clients to use, our web application programme will be installed on the computer. Because of this, it is practical for solo use.

1.3 SOFTWARE SPECIFICATIONS

1.4.1 Jupyter Notebook:

A web-based open-source programme called Jupyter Notebook is used for editing, producing live code, executing documents, and sharing documents with text, equations, and visualisations. Julia, R, and Python are the primary programming languages that are supported. Programmers can write Python programmes using the IPython kernel that is included with Jupyter Notebook. Other than IPython, there are more than 100 kernels that can be used.

1.4.2 Atom Text editor

All operating systems are compatible with Atom, a text and source code editor. It is a desktop programme created using the integration of HTML, JavaScript, CSS, and Node.js that speeds up find-and-replace operations by an order of magnitude and enhances loading performance for huge, single-line files.
Atom is a text editor that is free and open-source and made for coding. It is based on the Electron framework and was created by GitHub. The operating systems Windows, macOS, and Linux all support Atom, which also supports a large number of programming languages.

Because of its extensive customization and extensibility, Atom lets users customise the editor to suit their own requirements. With the use of themes and packages, its user interface is adaptable and may be customised. An integrated package manager in Atom makes it simple for users to set up and maintain packages that give the editor extra features like syntax highlighting for new programming languages or interface with version control systems like Git.

Atom's support for multiple panes, which enables users to view and edit numerous files simultaneously, is one of its important features. When working on tasks that require many files or comparing various versions of a file, this is especially helpful. Additionally, Atom has a built-in file tree view that facilitates file navigation and opening.

The ability to do typical Git actions, including committing changes, merging branches, and pushing and pulling changes, without leaving the editor is another helpful feature of Atom. This can lessen context switching between various tools and streamline the development process.

Atom is a potent coding tool thanks to its extensive range of integrated capabilities. It features assistance for autocomplete, a tool that suggests code snippets and function names as you type and helps speed up development. It also has an integrated debugger that enables users to step through code and troubleshoot issues.

Along with its built-in capabilities, Atom has a sizable and vibrant developer

community that actively contributes to its growth and produces new packages and themes. Users may now access packages that increase functionality or offer customization choices that suit their unique needs with ease.

Overall, Atom is a text editor built specifically for coding that is strong and incredibly flexible. It is a versatile tool for developers of all skill levels thanks to its support for multiple panes, Git integration, and built-in functionality. As a result of its extensibility and vibrant community, it is a well-liked option among developers who wish to customise their editor to suit their unique requirements.

Split Panes: The editor in Atom can be divided into several panes either horizontally or vertically. This capability is especially helpful when comparing code between files or while editing numerous files simultaneously.

Snippets: Atom comes with a built-in snippets feature that enables users to generate unique code snippets that can be quickly and easily put into their code. This function automates repeated coding processes, which can save time and lower errors.

Atom includes a robust find and replace capability that enables users to look for particular code snippets or strings within a file or across many files. Regular expressions can be used by users to carry out sophisticated search and replace operations.

Git Diff: Atom includes a Git Diff functionality that enables users to see changes made to a file in relation to a previous version. This functionality can make it simpler for developers to diagnose problems and understand the changes made to a file.

Code folding is a feature of Atom that makes it simpler to navigate through lengthy code files. This function is especially beneficial for huge files that contain numerous routines or nested loops.

Project Management: Atom includes a project management functionality that enables users to manage their coding projects right inside the editor. Users have the ability to start new projects, add files and directories, and arrange files and folders inside of projects.

Overall, Atom is a robust and flexible text editor created for programmers. It is a well-liked solution among developers of all experience levels due to its flexibility, customisation possibilities, and built-in capabilities. Atom may make writing code more productive and efficient, whether you're a new or seasoned developer.

1.4.3 SVM

Both classification and regression are performed using supervised machine learning techniques known as Support Vector Machines (SVM). The most relevant phrase is categorization, even though we also mention worries about regression. The SVM method aims to find a hyperplane in an N-dimensional space that clearly classifies the

data points. The size of the hyperplane depends on the number of features.

If there are only two input features, the hyperplane effectively looks like a line. If there are three input features, the hyperplane transforms into a 2-D plane. It becomes difficult to imagine something having more than three features. The SVM kernel is a function that takes a low-dimensional input space and transforms it into a higher-dimensional space, allowing it to solve problems that are not separable. For non-linear separation problems, it works best. Simply said, after completing some highly complex data transformations, the kernel decides how to split the data based on the labels or outputs defined.

SVM operates by locating the hyperplane in the feature space that best divides the two classes. The hyperplane is a line that divides the two classes in a two-dimensional feature space. The hyperplane is a hyperplane that divides the classes in a high-dimensional feature space. Finding the hyperplane that maximises the margin between the classes is the goal of the SVM algorithm. The margin is the separation between the nearest data points from each class and the hyperplane.

Both linear and non-linear classification can be done using SVM. The hyperplane is a linear function of the input features in the case of linear classification. The hyperplane is a non-linear function of the input features in non-linear classification, though. SVM uses the kernel trick to transform the input characteristics into a higher dimensional space where the classes are separated by a hyperplane in order to execute non-linear classification.

In comparison to other machine learning algorithms, the SVM method has a number of benefits. Its ability to solve classification issues that are both linear and non-linear is one of its main features. SVM also focuses on locating the hyperplane that maximises the margin between the classes because it is a margin-based algorithm. As a result, the model is less susceptible to data noise and outliers.

SVM has certain restrictions as well. The biggest drawback of SVM is that it can be vulnerable to the selection of a kernel operation. Finding the best kernel function is frequently a matter of trial and error. The choice of kernel function can have a considerable impact on the performance of the SVM algorithm. When working with huge datasets, SVM can also be computationally expensive, especially when employing non-linear kernels.

In conclusion, SVM is a potent machine learning technique that may be utilised for regression and classification. It operates by identifying the hyperplane in the feature space that best divides the classes. SVM is less sensitive to noise and outliers in the data and is capable of handling both linear and non-linear classification problems. When working with huge datasets, it might be computationally expensive and dependent on the kernel function selected.

The following are some of the salient characteristics of SVM:

Kernel Trick: SVM uses a method known as the kernel trick to handle non-linearly separable data. The input data are changed using the kernel trick into a higher-dimensional space where a hyperplane can be used to divide them.

Margin Maximisation: SVM identifies the hyperplane between the two classes that maximises the margin between them. The margin is the separation between the nearest data points from each class and the hyperplane. By increasing the margin, one can make the classifier more resistant to noise and outliers in the data by giving it a wide buffer zone between the two classes.

Support Vectors: SVM only use support vectors, a subset of the training data, to identify the hyperplane. The data points that are closest to the hyperplane and have the biggest impact on where it is are known as support vectors.

SVM use regularisation to reduce overfitting and enhance generalisation. performance. Regularisation increases the objective function's penalty term, which incentivizes the classifier to have a less complicated decision boundary.

1.4.4 CNN

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning technique that can take an input image, assign different elements and objects in the image importance (learnable weights and biases), and be able to differentiate between them. The use of CNNs for image categorization and recognition is made possible by their high degree of accuracy. It was first proposed by computer scientist Yann LeCun in the late 1990s when he became interested in how people recognised objects visually. Image processing, classification, segmentation, and other auto correlated data are the main uses of a convolutional neural network (CNN), which consists of one or more convolutional layers. They have the capacity to automatically learn how to extract hierarchical representations of picture information thanks to their deep learning architecture.

Many different applications, including image classification, object identification, face recognition, and natural language processing, make extensive use of CNNs.

The visual cortex in the human brain serves as an inspiration for CNNs. They are made up of several layers, each of which applies a particular operation to the input data. The fundamental component of a

The convolutional layer, or CNN, converts the input image into a set of learnable filters. Each filter creates a feature map that draws attention to a particular pattern or aspect of the input image, such as textures, edges, or corners.

The rectified linear unit (ReLU), which provides non-linearity into the model and enables it to learn more complex representations of the input features, is then fed via the output of the convolutional layer. The pooling layer decreases the spatial size of the feature maps while keeping their most crucial features by passing the output of the activation function through it.

One or more fully linked layers that can learn to categorise the incoming data into different groups come after the pooling layer. A prediction for the input image is generated by the fully connected layers using the flattened output of the preceding layers and a set of weights applied to each feature.

In comparison to other machine learning algorithms, CNNs have a number of benefits. Their ability to automatically learn how to extract hierarchical representations of picture characteristics without the help of a human being is one of their main advantages. Furthermore, CNNs are capable of handling big and intricate image datasets, and they generalise effectively to fresh, untried images.

CNNs do have some restrictions, though. The fact that CNNs need a lot of training data

in order to reach high accuracy is one of their key drawbacks. Furthermore, CNNs can be computationally costly, particularly when working with high-resolution images. For them to work well, hyperparameters like the number of layers, the size of the filters, and the learning rate need to be carefully tuned.

Here are some of the advantages of CNNs:

1. Robust to Translation: CNNs are capable of recognizing objects even if they are translated or rotated in the image.

2. Hierarchical Representation: CNNs learn hierarchical representations of the image, with each layer of the network learning increasingly complex features.

3. Automatic Feature Extraction: CNNs automatically learn the features that are most important for the classification task, rather than requiring hand-engineering of features.

4. High Accuracy: CNNs are currently the state-of-the-art in many image recognition and classification tasks.

However, CNNs also have some limitations:

1. Large Number of Parameters: CNNs can have a large number of parameters, particularly for large images. This can make training the network computationally expensive and time-consuming.

2. Limited Interpretability: Because CNNs automatically learn features, it can be difficult to interpret why the network is making a particular classification decision.

3. Overall, CNNs are a powerful and effective technique for image recognition and classification tasks. Their ability to automatically learn hierarchical representations of the input data, combined with their robustness to translation, make them a popular choice in many computer vision applications.

In conclusion, CNNs are a potent deep learning architecture that has the ability to automatically learn to extract hierarchical representations of picture information. They can manage huge and complex datasets and are frequently employed in image and video identification tasks. However, they might be computationally expensive and call for a lot of training data

2. LITERATURE REVIEW

Human trafficking is one of the most challenging problems for law enforcement to address and calls for international cooperation. In order to identify the most likely advertising that are associated to human trafficking and identify potential patterns of online human trafficking operations, we used readily available data from the classified ad website "Backpage" in this study. Because there is no ground truth, we rely on two human analysts—one who has survived being a victim of human trafficking and one from law enforcement—to manually classify the scant crawling data. Then, we offer a semi-supervised learning technique, which is evaluated on unobserved data with additional expert verification and trained on the available labelled and unlabelled data.

Existing System Summary

Although there have been prior attempts at picture classification and tweet filtering to identify unlawful communications, the majority of these systems rely on computer vision or natural language processing. The survey used semi-structured questionnaires to gather data on the victims' use of technology both before and after they were trafficked, how they advertised themselves, the various services and technologies that were used to trade in trafficked people who were sexually exploited, and how clients looked for, communicated with. The research results showed that traffickers and their networks used sophisticated software to safeguard their anonymity, internet hosting and storage services, and cutting-edge encryption techniques to frustrate police digital forensic examinations. Because our goal is to identify the invariant characteristics of deception in text, we propose that these encouraging results in automatic deception detection are mostly the result of corpus-specific features. This does not prevent practical applications, but it does not promote a deeper investigation of dishonesty. To demonstrate this and allow researchers and practitioners to exchange findings, we developed the BLT-C (Boulder Lies and Truths Corpus), the biggest publicly accessible shared multidimensional deception corpus for online reviews. This thesis highlights how challenging it is to detect deception using supervised machine learning approaches. To overcome the inherent absence of ground truth, we have also developed a series of semi-automatic algorithms to guarantee corpus validity. This global problem, which robs millions of victims of human trafficking of their dignity, causes them to suffer. Currently, social networks are used to spread this criminal activity online by broadcasting covert messages that promote these illegal services. Due of the limited resources available to law enforcement, it is critical in this case to automatically find texts that may be connected to this crime and may serve as clues. Text messages are more common than ever, and text-based false language is increasingly used in cybercrime. We use machine learning and linguistic approaches to detect deceit in text exchanges in cybercriminal networks. We develop cybercrime detection methods using web genre. Our contributions include the following: models trained on email fraud can predict social media fraud with 50% predictive accuracy; models trained on social media fraud can detect email fraud with 60% predictive accuracy. The outlook for the email model is favourable because the hackers in this study speak a variety of languages. We also demonstrate how cybercrime detection models may be constructed using components of natural language processing and linguistic psychological processes connected to cybercrime.

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)	-FFFN - CNN	- SVM -CNN -Lasso
Evaluation parameters	- Model Accuracy	- Model Accuracy	- Model Accuracy

Table 2.2: Literature review summary

2.1 Proposed System

Using SVM and Naive Bayes machine learning techniques, the author of this study proposes a concept to detect human trafficking through the analysis of social media text conversations. The author of this research initially searches Twitter using terms like Lolita, escort, and others. The extracted tweets are then cleaned to remove special characters and stop words (such as the, where, and, an, etc.), and they are then analysed to extract words like VERBS and ADJECTIVE, which may contain crucial information or suspicious words used by HUMAN TRAFFICKERS (such as chicken soup, girls, penguin, and more).

Clean tweets will be sent to SVM and Nave Bayes classifiers as input to identify suspicious phrases.

3. PROBLEM FORMULATION

Online Twitter Crawl: This module lets us enter a hashtag and have the programme search Twitter using the TWEEPY API for tweets that contain that term.

If you don't want to crawl Twitter, you can upload an existing Twitter dataset using the offline upload feature of this module.

This module's "Clean Tweets & Extract Features" removes special characters and stop words from each tweet before separating its verbs and adjectives, which are then put into the SVM and Naive Bayes algorithms. SVM performs better than the Naive Bayes method at identifying suspicious tweets.

In this module, we feed clean tweets to the SVM and Naive Bayes algorithms for the classification of suspicious tweets. The entire data set is then split into train and test sections, with 80% of the data being used for training and 20% being used for testing. 80% of the data will be used to construct a model after algorithms have been trained. A trained model will be used to evaluate prediction accuracy, precision, recall, and FSCORE using test data.

Following the detection of suspicious tweets, each suspicious tweet website is scanned to read all photos, and from those images, the face and upper body components are extracted using SVM classifier. SVM & CNN Classification for Gender & Age Prediction.

Every machine learning effort needs to start with the formulation of problems. It include describing the issue that needs to be resolved, locating pertinent data sources, and choosing the best techniques and algorithms for data analysis. Natural language processing (NLP) and computer vision techniques can be used to identify potentially dangerous content on Twitter, but we must carefully frame the problem in order to effectively identify and highlight potentially harmful content.

For this task, the problem formulation procedure is described in greater detail below:

Clearly defining the issue that needs to be solved is the first stage in problem formulation.

In this scenario, our objective is to create a system that can identify potentially illegal Twitter messages. Messages including hate speech, cyberbullying, the sale of drugs, or other unlawful actions may fall under this category. Twitter should be constantly scanned by our system for potentially hazardous content, which will then be flagged for further inspection by moderators.

Finding the appropriate data sources for our project is the following phase in the data collection process. In this situation, a dataset of tweets with known illegal content must be gathered. This can entail looking for tweets that have been reported as offensive or utilising terms that are frequently connected to illegal content. To make sure that our model is not biassed, we need also take into account any potential biases in the dataset and take action to address them.

Data preprocessing: Once the dataset has been gathered, the data needs to be prepared for analysis. Stop words, punctuation, and other distracting elements from the text must be eliminated, and the data must then be further refined using stemming, lemmatization, and tokenization methods. Additionally, we must use computer vision algorithms to preprocess any pictures or videos that are posted on Twitter.

Natural Language Processing (NLP): Using the preprocessed data as a starting point, NLP techniques can be used to recognise and classify tweets. This might entail employing named entity recognition to find significant entities like people, locations, and organisations, sentiment analysis to determine the emotional tone of the tweet, topic modelling to categorise tweets into themes and spot trends in the data, and sentiment analysis to determine the emotional tone of the tweet.

Computer Vision: Using computer vision techniques, we can examine both the text and the photos and videos uploaded on Twitter. Using machine learning algorithms, this entails classifying pictures or movies as possibly illegal by spotting patterns in the visual data. For instance, we can utilise object recognition to find pictures of criminal activity, guns, or drugs.

Machine Learning: We may use a machine learning model to categorise tweets as possibly unlawful or not after utilising NLP and computer vision techniques to extract features from the data and preprocess it. This could entail identifying trends in the data and making assumptions about whether a specific tweet is appropriate or not using methods like logistic regression, decision trees, or neural networks.

Evaluation: After our machine learning model has been trained, we must assess its effectiveness in spotting possibly illegal texts. This could entail evaluating the model's accuracy using measures like precision, recall, score as well as performing a human evaluation of the flagged tweets to determine whether or not they are actually unsuitable.

Finally, we must put our system into operation and integrate it with a monitoring tool that can continuously search Twitter for potentially hazardous content. Building a web application that enables human moderators to examine flagged tweets and take appropriate action, like deleting the tweet or banning the user who tweeted it, is one way to accomplish

this.

In conclusion, the problem formulation process for identifying potentially illicit tweets on Twitter using NLP and computer vision techniques entails defining the issue clearly, gathering and preprocessing the pertinent data, using NLP and computer vision to extract features from the data,

4. OBJECTIVES

The suggested study is intended to be carried out in order to build a strategy for stopping human trafficking. The system, which is the subject of the proposed work, will be accomplished by breaking it down into the following goals:

- Access the Big Maty dataset online.
- Wrangle data by yourself.
- Develop various machine learning algorithms.
- To reach the highest level of accuracy, try to create algorithms based on various models.

To use CNN, a Deep Learning Algorithm, to deploy the model once it has been developed using SVM.

Twitter is a well-known social networking site that enables users to send and receive brief messages (called tweets). While the majority of tweets are innocent, some may include offensive language, cyberbullying, the sale of illegal substances, or other unlawful activity. Such tweets can be identified and blocked to keep Twitter secure and protect its users.

We can analyse tweets and find possibly illegal messages using a combination of NLP and machine vision techniques to solve this issue. The steps in this technique are as follows:

Data collection: We require a dataset of tweets that are known to include illegal messages in order to train a machine learning model to recognise such comments on Twitter. This dataset can be gathered by looking for tweets that have been reported as offensive or by utilising keywords that are frequently used in connection with illegal content.

Data preprocessing: After gathering the dataset, the data needs to be prepared in order to be more easily manipulated. Stop words (frequent words with little significance), punctuation, and other distracting elements from the text must be eliminated in order to

accomplish this. To further hone the data, we can additionally employ strategies like stemming (reducing words to their simplest forms), lemmatization (changing words into their dictionary forms), and tokenization (dividing text into individual words).

After preprocessing the data, we may use Natural Language Processing (NLP) methods to identify and classify the tweets. The emotional tone of the tweet can be determined by sentiment analysis, and significant entities like people, locations, and organisations can be found using named entity recognition. Additionally, topic modelling can be used to categorise tweets into themes and spot trends in the data.

Computer Vision: Using computer vision techniques, we can examine both the text and the photos and videos uploaded on Twitter. Using machine learning algorithms, this entails classifying pictures or movies as possibly illegal by spotting patterns in the visual data. For instance, we can utilise object recognition to find pictures of criminal activity, guns, or drugs.

Machine Learning: We may use a machine learning model to categorise tweets as possibly unlawful or not after utilising NLP and computer vision techniques to extract features from the data and preprocess it. Using the preprocessed data and the extracted features as input, a model is trained in this manner. The fresh tweets that are sent on Twitter can then be categorised using the model.

Integration: Lastly, we can combine a monitoring system that continuously scans Twitter for possibly illegal statements with the machine learning model. Human moderators can check any tweets that have been flagged as possibly illegal to see whether they are in fact inappropriate or not.

In conclusion, combining NLP and computer vision methods can aid in the detection of potentially illegal posts on Twitter and help to maintain the service secure for its users. However, it is crucial to keep in mind that any system created to keep track of social media activity must adhere to privacy laws and regulations and be regularly updated and improved to take into account new trends and modifications in the kinds of messages broadcast on the platform.

5. METHODOLOGY

The objectives outlined for the suggested research activity will be accomplished using the technique described below:

- The topic of human trafficking will be thoroughly researched.
- There will be installation and practical application of current countermeasures against human trafficking. The relative benefits and drawbacks will be noted.

- The suggested system will be evaluated using a number of parameters.

- There will be a comparison between the newly adopted strategy and the current approaches.

Natural language processing (NLP) and computer vision methods are used in a multi-step procedure to find potentially illegal statements on Twitter. The process is explained in more depth below:

Data Gathering: The first phase entails gathering a dataset of tweets that are known to include offensive content. This can entail looking for tweets that have been reported as offensive or utilising terms that are frequently connected to illegal content. To make sure that our model is not biassed, we need also take into account any potential biases in the dataset and take action to address them.

Data preprocessing: Once the dataset has been gathered, the data needs to be prepared for analysis. Stop words, punctuation, and other distracting elements from the text must be eliminated, and the data must then be further refined using stemming, lemmatization, and tokenization methods. Additionally, we must use computer vision algorithms to preprocess any pictures or videos that are posted on Twitter.

Natural Language Processing (NLP): Using the preprocessed data as a starting point, NLP techniques can be used to recognise and classify tweets. This might entail employing named entity recognition to find significant entities like people, locations, and organisations, sentiment analysis to determine the emotional tone of the tweet, topic modelling to categorise tweets into themes and spot trends in the data, and sentiment analysis to determine the emotional tone of the tweet.

Computer Vision: Using computer vision techniques, we can examine both the text and the photos and videos uploaded on Twitter. Using machine learning algorithms, this entails classifying pictures or movies as possibly illegal by spotting patterns in the visual data. For instance, we can utilise object recognition to find pictures of criminal activity, guns, or drugs.

Feature Extraction: After preprocessing the data and utilising NLP and computer vision techniques to extract features, we can utilise these features to train a machine learning model to categorise tweets as possibly unlawful or not. This could entail identifying trends in the data and making assumptions about whether a specific tweet is appropriate or not using methods like logistic regression, decision trees, or neural networks.

Model Training: Using the preprocessed data and the machine learning algorithm we choose after extracting the features. In this case, the data would be divided into training and testing sets, with the training set being used to train the model and the testing set to assess how well it performed.

Model Evaluation: After our machine learning model has been trained, we must assess its effectiveness in identifying potentially illegal texts. This could entail evaluating the

model's accuracy using measures like precision, recall, and F1 score as well as performing a human evaluation of the flagged tweets to determine whether or not they are actually unsuitable.

Finally, we must put our system into operation and integrate it with a monitoring tool that can continuously search Twitter for potentially hazardous content. Building a web application that enables human moderators to examine flagged tweets and take appropriate action, like deleting the tweet or banning the user who tweeted it, is one way to accomplish this.

Overall, a machine learning model is trained to categorise tweets as potentially illicit or not, its performance is assessed, and the system is then deployed for continuous monitoring of Twitter. This methodology for detecting possible illicit messages on Twitter using NLP and computer vision techniques involves collecting and preprocessing the relevant data, using NLP and computer vision to extract features from the data, and training a machine learning model to classify tweets as potentially illicit or not.

6. 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.

.

CHANDIGARH UNIVERSITY
Discover. Learn. Empower.