An article

on

Application of Text Mining in Identifying Illicit Supply Chain Network

Submitted to:

Dr. Mahdi Fathi

Submitted by:

DSCI5240-Group 11

Mahima Yalavarthi 11522517

Sabnam Shrestha 11446896

Yamini Mangarai 11535594

November 14, 2021

Illicit Supply Chain is getting more pervasive with the increased access of technology to all. Especially, the anonymity that social media allows has made it easier for the Illicit Supply Chain networks to run covertly, and to reach common people and turn them into their targets.

Luckily for us the increased access to data and data mining algorithms means that we can detect such network operations in Social-Media with the help of classification techniques like Text Mining. In which we can select the ‘Keywords’ that are likely to be used by people operating such networks and track whether certain social media activities such as posts and events or social media account presence are operated by the potential drug dealers.   
  
We are especially interested in identifying such potential social media drug dealing networks and dark web services, as they are more likely to freely access the data about the personal details, demographics, likes and dislikes of youths and then target them into selling such drugs and even coercing them to be a part of such networks, turning it into a vicious cycle.

In data mining, we can use classification technique such as the Naïve Bayes Algorithm, which uses conditional probability, joint probability, proportionality and bayes theorem that are applicable in analyzing discrete variables and determining the probability of an event based on the knowledge of conditions as inputs variables related to the probability of outcome of a certain event.

Texts are ubiquitous in Social Media platforms. The people who are involved in illicit drug smuggling make use of texts through messaging, hashtags, imagery and posts, using certain codewords specific to their chains and network. These codewords can include code names for illicit drugs, slang terms, location address codes and can indicate the potential illicit drug supply conversations and threads in the web.

We can train our algorithm by including an existing set of codewords that are known to indicate illicit drug operation and categorize them into classes; then predict ‘Keywords’ and ‘Non-Keywords’ that are identified with the help of the word classes as to whether these words represent a social media message or post for drug operation or not.

We can also prepare an evaluation matrix as follows:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Username | Weed | White Powder | Pot | Meth | Bruh |
| Random@1 | 0 | 0 | 1 | 0 | 1 |
| Random@2 | 1 | 0 | 1 | 0 | 0 |
| Random@3 | 0 | 1 | 1 | 1 | 0 |
| Random@4 | 0 | 0 | 0 | 0 | 1 |

Then, we can calculate the probability of the chances of being a Keyword given the frequencies of these words in a certain text. Where, X is the combination of the words present in Evaluation matrix and y is the prediction of being a ‘Keyword’ used by drug smugglers.



Text Mining is a useful tool for identifying and predicting whether a certain set of words are related to a certain outcome. This can help us in extracting interesting patterns that are hidden from plain sight. Specifically, Social Media platforms are easy targets for illicit drug supply operations, thus, having the ability to process such texts in real time and taking necessary steps can enhance online safety for youths and tackle problems related to drug addiction by preventing smugglers from targeting teens, youths and other innocent individuals using social media. We can also add a recommendation system or alert based on our analysis to send warning messages of being a potential illicit site or an inauthentic user by further applying Text Mining and subsequent steps.

It has become increasingly important for the highest officials of popular Social Media Platforms like Instagram, Facebook, Twitter and Tiktok to be vigilant of the content shared on these platforms. They must designate proper policy and guidelines in place to actively detect the misuse of such sites at the hands of drug peddlers and smugglers. Statistically, in a study of 358 online participants for “#Drugsforsale: An exploration of the use of social media and encrypted messaging apps to supply and access drugs,” 76 percent said they used Snapchat to acquire drugs and 21 percent told they accessed drugs through Instagram. Likewise, looking at the supplier side, many pharmacies exist online nowadays, however, there is a blurred line between legitimate versus illegitimate pharmacy businesses, a review by the National Association of Boards of Pharmacy (NABP) found that 97% of the pharmacy websites violate pharmacy laws and practice standards. Without gatekeepers for online marketplace and online businesses social media sites have become more vulnerable to illicit supply chain.

In essence, the growing popularity of social media is a double-edged sword, at one hand legitimate businesses can reach and serve their target customers spread across the world, but at the other end of spectrum, dark web and social media drug networks are operating under cover of harmless businesses. With the help of data mining techniques like text mining, sentiment analysis, and recommendation systems, we can identify and disrupt the communication networks for such illicit drug suppliers and ensure the safety and protection of social media users. This should be a concern not just for social media platform owners and stakeholders but, for the data mining community, the CIOs, general public, parents and teachers, to be aware of what kind of content is being circulated in social media. Surely, keeping policies and algorithms to detect and disrupt such illicit networks can be a place to start with in this e-fight against illicit supply chain networks.

References:

<https://towardsdatascience.com/a-mathematical-explanation-of-naive-bayes-in-5-minutes-44adebcdb5f8>

<https://towardsdatascience.com/implementing-a-naive-bayes-classifier-for-text-categorization-in-five-steps-f9192cdd54c3>

<https://americanaddictioncenters.org/drugs-on-social-media>

<https://novarecoverycenter.com/drug-use/social-media-contributes-drug-abuse/>

Social Media Text Mining Framework for Drug Abuse: An Opioid Crisis Case Analysis by Tareq Nasralah Dakota State University