

Detecting Illegal Drug Ads and Suppliers across Social Media

M R Sumalatha¹, R Deekshetha², R S Vineetha³ and Rasmia Rahamathullah⁴

¹⁻⁴Department of Information Technology, Anna University (MIT Campus), Chennai, India
Email: sumalatha@annauniv.edu

Abstract—Teenagers and young people today spend a lot of time on social media. The youth who use social media have a higher chance of using alcohol, tobacco, and illegal drugs due to various sources available online. Teens' risk of substance use and addiction may be reduced by limiting their exposure to social media posts on illegal drug use and promotion. These platforms have become a source for buying and selling illicit drugs online. Colloquial language used to caption the images associated with drugs is one of the challenging issues for filtering out such posts on social media. In this paper, we offer a technique for automatically identifying social media posts related to illicit drug promotion. With the help of this technique, social media may automatically filter out anything that is associated with illicit drugs. The proposed model uses state-of-the-art social media analytics, which combines text and image processing based on neural networks, to find posts related to illicit drug promotion on social media. Bert tokenizer is used to extract textual features and VGG16 is used for feature extraction from images. Evaluation results show that textual features produced from word embedding and image features derived from VGG16 neural network, when combined together shows better accuracy in classifying posts related to illicit drug promotion compared to other statistical models.

Index Terms— Social media, Illicit drugs, Embedding, VGG16, BERT tokenizer, Neural Networks, Multimodal.

I. INTRODUCTION

With the growth of social media and the expansion in the use of smartphones, a dangerous threat has emerged among youngsters. Criminal drug networks are abusing social media to expand their reach, create new markets, and target new clients. This includes selling deadly opioids and LSD pills, often to unsuspecting teenagers, young adults illegally. Selling and buying of these lethal drugs are no longer confined to the dark web and criminal drug networks, it is now easily available in every home because of internet apps on our smartphones. Increased drug accessibility and promotion made possible by the Internet is one factor fueling the epidemic. Vendors frequently use social media platforms to advertise illicit drugs because they are relatively simple to use, have a wide audience, and are hard to detect the trafficking.

Various social media platforms like Twitter, Instagram use many algorithms to stop illicit drug promotion advertisement posts to prevent drug abuse through these social media. But illegal drug dealers somehow find ways to evade the previously set up restrictions by using slang words which these social media platforms fail to detect. Vulnerable young teenagers could fall prey to these illegal activities and end up risking their lives. Many of this may go unnoticed by parents and hence it becomes crucial to find new ways to better detect these

activities online.

Finding active vendors by identifying vendor groups in social network is a new surveillance method to prevent drug sale online. It is seen that drug vendors have connections in social media to promote their drugs online among their self-formed community. In the same community, there might be active relationships between suppliers, which could indicate trends in their corresponding networks and easier to locate much more illegal drug selling vendors. The proposed model analyses posts on social media platforms for advertisements and promotions of illicit drugs and connects supplier's using community detection.

II. RELATED WORKS

A. Tweet Classification

Ryan Sequeira et al. performed a large-scale analysis on the spreading of Drug abuse tweets on twitter[1]. The study reveals the existence of a very large giant component involving 99% of the users with dense local connectivity that facilitates the spreading of drug abuse messages. Furthermore, the cascades of Drug abuse tweets gets spread over a long distance through the engagement of several closely connected groups of users. In their work, Naive Bayes, SVM, Random Forest, Logistic Regression were implemented and it was observed that SVM trained outperformed all the classifiers with a combination of sentence embedding.

The research demonstrated by Tao Ding[4] shows how interpretable behaviour with high prediction performance can be built with a knowledge distillation framework. This method has been applied to social media based substance abuse prediction. The evaluation results shows good prediction performance and it is interpretable. The knowledge from the models gives insights into the relationship between social media activities and substance abuse.

A deep learning based approach consisting of CNN and LSTM neural architectures using WordNet and ConceptNet to improve the coverage in Embedding space and provide better initialization for further deep learning architecture is developed by Gauri Jain[7]. The model is supported by introducing the semantic information in representation of the words with the help of knowledge-bases such as WordNet and ConceptNet. Use of these knowledge bases improves the performance by providing better semantic vector representation of testing words.

In the research paper by N. Phan et Al. [8], the primary goal was to demonstrate the possibility of utilizing social media, e.g., Twitter, for automatic monitoring of illegal drug and prescription medication abuse. They used machine learning methods for an automatic classification that can identify tweets that are indicative of drug abuse. Their experiment compares a set of classification algorithms and a decision tree classifier J48, and observed that SVM outperformed others with a precision of up to 74.8%.

A. Kok in his conference paper[9] addresses the problem of how to achieve similarity search for text sequence correlation with a proper semantic foundation. It presents and compares three methods for similarity search. The method with higher performance were implemented using Google introduced transformer model, BERT.

Tim K. Mackey et al. [12], uses a Bitern Topic Model (BTM) Unsupervised Machine Learning to detect and extract tweets associated with online marketing and sale of controlled substances. It filters a large volumes of Twitter content to specifically isolate tweets promoting the illegal online sale of controlled substances, an activity contributing to the current national prescription drug abuse epidemic.

B. Community Detection

Zhang Feng Yin et al. in [2] used contact graphs and user social behaviour to link nodes, which is capable of identifying not only the set of adjacent nodes that belong to one user but also the set of non adjacent IDs that belong to the same user using User Identify Set Matching Problem (UISMP) Algorithm. The paper utilizes a set matching algorithm based on the contact graph and the built universal model to dispose the set of candidate node IDs, and select the most appropriate match in line with confidence score gained by means of this algorithm.

Chao Han proposed a Tensor-based Markov chain (T-Mark) approach[5], which is able to automatically and simultaneously predict the labels for unlabeled nodes and give the relative importance of types of links that actually improve the classification accuracy. Extensive experimental results demonstrate that the T Mark approach is able to achieve superior performance in the comparison and obtain reasonable relative importance of links. Wanqiu Cui in his paper[6] proposed that MVGAN mitigates the barrier of short text sparsity and enriches its expressiveness from two perspectives:

(1) A heterogeneous graph is built to associate the context information of nodes to each short text. The GCN is adopted to combine the attribute information and associated structure in each view, which enables the effective expansion of short texts by aggregation of neighbour nodes.

(2) And designed a multi-view graph attention mechanism to guide the fusion of semantic and temporal views by leveraging the hashtags as the bridge across different views. It mines the coherence between different view features and realizes the feature enhancements of short text.

A graph based approach to detect drug abuse is discussed by K.M. Saifuddin. A text graph is built at the corpus and document levels to capture document-word, document-document, and word-word relations using tweets about drugs. Additionally, a variety of GNN models were built on these text graphs, transforming the text classification task into a node classification task for corpus-level graphs and a graph classification task for document-level graphs, respectively, in order to identify DA events.

Based on the modularity of the vertex relations, K. Sutaria et. al. proposed a community detection technique [11]. A method based on dynamically calculating modularity was developed by them. At each phase of preprocessing, new vertex is added or refreshed with local area record. At the final stage, obtained vertex list with its community and modularity of partitions.

III. PROPOSED ARCHITECTURE AND SYSTEM DESIGN

The goal of the proposed architecture is to identify user profiles which promote and advertise drugs on social media platforms. For this purpose the proposed model will perform several levels of filtration on user profiles. Social media platforms have facilitated illicit drug trading and may be an important driver of the current opioids epidemic. Thus the proposed model aims to prevent the social platforms from being used as a trading means of illicit drugs. The three modules in the architecture have their significance in detecting the drug traffickers. The first module identifies those posts which are advertising drugs while the second module suspects other users who may be involved in the drug trafficking. The Multimodal identifies the drug dealers from the community detected in the second phase. It classifies the profiles based on the image and text data from tweets shared by the user. Initially Event detection is carried out to classify drug ads from other ordinary tweets. The event to be detected is the promotion or selling of illicit drugs.

Social media platforms have given its users the freedom to share information in the form of text and multimedia. These platforms have no control over the kind of information that is being posted online. Drug dealers make advantage of this situation to promote their drugs. To combat this issue, the event detection module performs the first layer of abstraction. The textual data from tweets are classified using two different neural network models, CNN-BiLSTM and BERT-LSTM model in identifying tweets that promote illicit drugs.

The profiles which were isolated in event detection is provided as input to community detection. Then, Community detection is performed where groups of users who are found to be promoting drugs are recognized. The follower list and following list of the identified profile may include other profiles that could also be involved in the selling of drugs online. To discover this group of drug vendors Louvain partition algorithm and Greedy modularity algorithm were employed. Finally, a Multimodal architect is deployed to efficiently classify profiles based on their most recent tweets. The top tweets of a particular profile is analyzed using this approach to categorize the twitter profiles into drug promotion, non drug promoting. The base models include VGG16 and Text embedding layer which perform feature extraction. The image tower and text tower are stacked on top of the respective base models to learn the extracted features. The final dense layer concatenates the two classification reports and gives the final result.

IV. EXPERIMENTAL ANALYSIS

A. Tweet Classification

Tweet classification is done to identify tweets that promote drugs from other educational drug tweets. Two models, CNN-BiLSTM and BERT-LSTM were compared to find the best model for the textual classification. CNN alone is capable of extracting features from a given input vector while BiLSTM alone is capable of identifying the patterns from the sentence vector. The input from Level 0 - embedding layer is passed onto the Level 1 - CNN layer where feature extraction is done. The extracted features are then fed to the Level 2 - BiLSTM layer where meaningful inference from the input vector is identified.

Similarly in BERT-LSTM model, a pretrained BERT encoder is used to extract the features from the sentences using two training strategies, Masked LM (MLM) and Next Sentence Prediction (NSP) by finding word's context depending all of its surroundings (left and right of the word). This extracted feature is fed into LSTM layer to classify the texts.

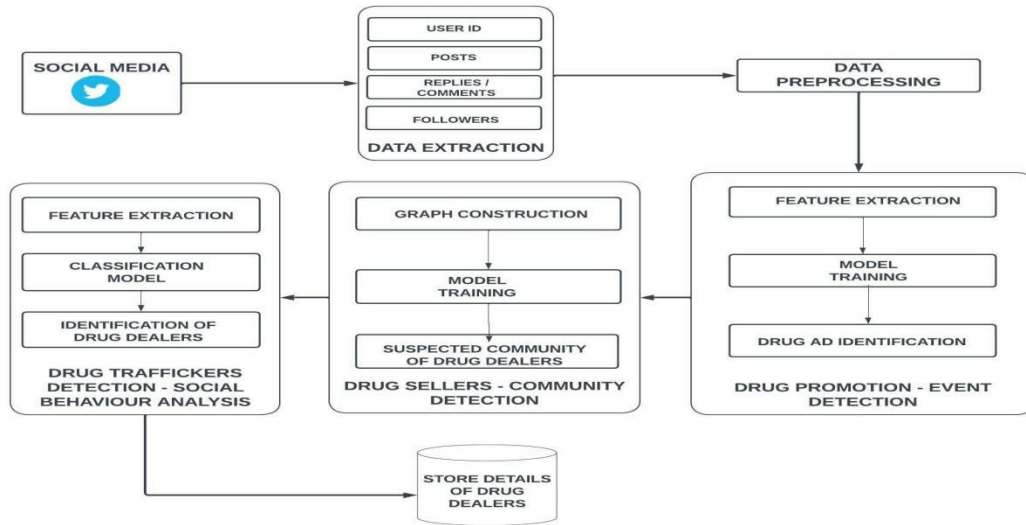


Figure 1. System Architecture

B. Community Detection

The Louvain algorithm is implemented which is an efficient hierarchical clustering algorithm based on graph theory. Its principle is to make the modularity of community partition results reach the maximum value through continuous iteration of mobile nodes and then obtain the optimal community partition.

Modularity(Q)

$$Q = 1/(2m) * \sum_{ij} ([A_{ij} - k_i k_j / (2m)] * \delta(c_i, c_j))$$

m = Weights of all edges in the graph

A_{ij} = weight of the edge between node 'i' & node 'j'

K_i is the degree (total connections) of some node i.

$\delta(c_i, c_j) = 1$ if both node i & node j are in same

C. Multimodal

Multimodal approach has been employed for identifying twitter profiles which may be selling drugs. It addresses some of the original goals of artificial intelligence by integrating and modelling multiple communicative modalities, including linguistic, and visual messages. The twitter dataset is a combination of text with images i.e , the tweets posted with media content - the image and corresponding caption. The meaningful features from text data and image data were extracted using BERT tokenizer and VGG16 respectively. And the two modalities of data , image using image tower model and image caption using text tower model were trained in parallel. Finally, both the models were combined and final layers were added for classification of user profile.

V. PERFORMANCE AND RESULTS

A. Tweet classification

The drug related posts were collected by searching selective keywords related to banned illegal drugs. Specifically, colloquial slang of illegal drugs were used for the tweet search to get better results on illegal drug promotion ads. The annotated dataset were fed into existing and our proposed model. From the below graphs, we can conclude that the previously existing models like CNN, LSTM performed poorly in comparison to our models. While CNN had an accuracy of 77.6%, the CNN Bi-LSTM showed 80.75%. Our models BERT-LSTM gave 75% accuracy.

The combined CNN-BiLSTM model gives good results over the long text, since it benefits from the CNN's ability to extract features and the BiLSTM's characteristic to learn long-term bidirectional dependencies of the text.

B. Community Detection

The performance of the Louvain partition algorithm that partitions nodes based on their relationship between user_id, likes, comments and retweets is calculated using the modularity. Modularity is calculated as value

TABLE I. ACCURACY OF IMPLEMENTED MODELS

MODELS	ACCURACY
CNN	77.61
LSTM	75.98
CNN-Bi LSTM	80.75
BERT-LSTM	75

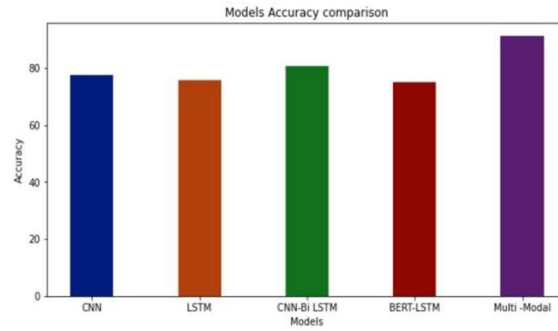


Figure 2. Graphical representation of model accuracy

between -0.5 to 1 i.e, Non-modular clustering to Fully modular clustering. It gives the measure of the relative density of edges inside communities with respect to edges outside communities.

The Louvain partition of a chosen drug dealer's two step followers list is implemented in figure 3. The modularity (Accuracy of community partition) is calculated using weights which includes number of retweets, likes and comments of each follower to the drug dealer posts. The following figure 3 visualizes 5 important communities for the drug dealer account.

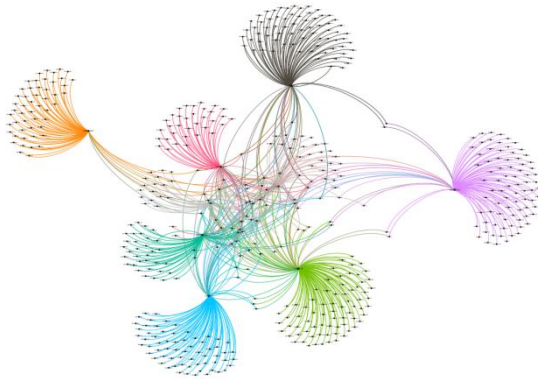


Figure 3. Partition found using community detection

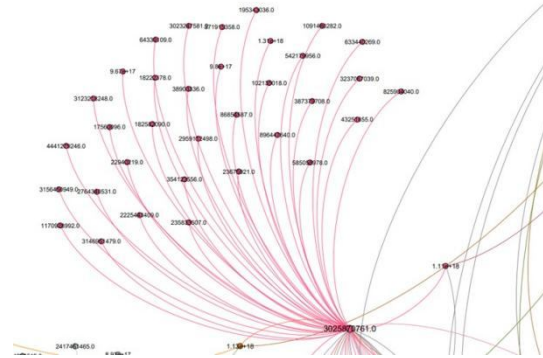


Figure 4. Single node connection in the community

C. Multimodal

The Multimodal implemented contains two parallel tower models, one for text feature extraction and another for the image feature extraction. The text features were extracted using the BERT tokenizer and the image feature is retrieved using VGG16. In the output layer, concatenation of the two models- text tower and image tower is performed. Soft max activation is used as this is a classification algorithm. The metrics used in the our experiment is cross modal gap. The embeddings from different modalities have different distributions, hence

cross modal gap makes universal embedding learning difficult. Since the images and text belong to the same instance, the ideal distance between them should be close to zero. So, these loss functions which are calculated to find the distance between the class predicted by image and text classification model is found to be 0.0018.

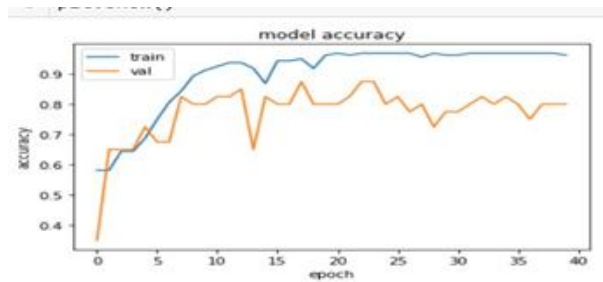


Figure 5. Training and Validation accuracy of Multimodal

VI. CONCLUSION

In this study, we used state-of-the-art picture and text analytics to find social media posts about illegal drugs. By removing posts about illegal drugs from social media, this technology can drastically lower the risk of exposing young adults to content on substance use. We gathered a Twitter dataset that was scraped to retrieve social media posts. In order to successfully detect posts related to illegal drugs, we employed both text features learnt using Embedding layers and picture features learned using Convolutional Neural Network as the input to a neural network-based classifier which shows an accuracy of 91% accuracy. Comparing this performance to models that just employ one kind of feature, it is noticeably superior.

Social media analytics that are text and image based are crucial to solving this issue completely. Our study serves as a significant first step in that approach. The proposed method can be further improved with more data and a robust feature representation technique. Our collection is considerably smaller than the ImageNet dataset (which has 1.2 million photos), despite the fact that it contains many scraped images. We anticipate that with more data, our proposed system accuracy can be increased significantly.

A social media post typically has brief, informal, and distracting text. For instance, there is a lot of flexibility in the spelling of a word on social media. Additionally, it's not unusual for social media messages to mix emojis and several languages. Grammar is also less important while blogging on social media. Traditional NLP methods like syntactic parsing are made more challenging by all of them. For text processing in this study, we use data-driven neural network-based machine learning approaches like Doc2Vec. But the amount of data used determines how well these operate. For instance, Doc2Vec struggled to generate accurate feature representations when the dataset was very small. We anticipate that by gathering more data, we will be able to significantly enhance system performance.

REFERENCES

- [1] R. Sequeira, A. Gayen, N. Ganguly, S. K. Dandapat and J. Chandra 2019, "A Large-Scale Study of the Twitter Follower Network to Characterize the Spread of Prescription Drug Abuse Tweets," in IEEE Transactions on Computational Social Systems, vol. 6, no. 6, pp. 1232-1244.
- [2] Z. Yin, Y. Yang and Y. Fang 2022, "Link User Identities Across Social Networks Based on Contact Graph and User Social Behavior," in IEEE Access, vol. 10, pp. 42432-42440.
- [3] Yongjian Ren , Yuliang Shi , Kun Zhang, Xinjun Wang , Zhiyong Chen, and Hui Li 2022, "A Drug Recommendation Model Based on Message Propagation and DDI Gating Mechanism, IEEE Journal of Biomedical and Health Informatics", Vol. 26, No.7.
- [4] Ding, T., Hasan, F., Bickel, W.K. and Pan, S., 2018. "Interpreting social media-based substance use prediction models with knowledge distillation. In 2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI) " (pp. 623-630) IEEE.
- [5] Chao Han, Jian Chen, Minghui Tan, Michael K. Ng, Qingyao Wu , "A Tensor-Based Markov Chain Model for Heterogeneous Information Network Collective Classification", IEEE Transactions on Knowledge and Data Engineering.
- [6] Wanqiu Cui, Junping Du, Dawei Wang, Feifei Kou, and Zhe Xue 2021, "MVGAN: Multi-View Graph Attention Network for Social Event Detection", ACM Trans. Intell. Syst. Technol. 12, 3, Article, 24 pages.
- [7] Gauri Jain , Manisha Sharma 2019, "Basant Agarwal, Spam detection in social media using convolutional and long short term memory neural networks", Springer Nature Switzerland AG.

- [8] N. Phan, S. A. Chun, M. Bhole and J. Geller, 2017 "Enabling Real-Time Drug Abuse Detection in Tweets," IEEE 33rd International Conference
- [9] A. Kok, G. Valiyev and M. Street 2021, "Adapting BERT Embeddings for Text Correlation of Military Domain Specific Content," 2021 International Conference on Military Communication and Information Systems (ICMCIS), pp. 1
- [10] Khaled Mohammed Saifuddin, Muhammad Ifte Khairul Islam, Esra Akbas 2021, "Drug Abuse Detection in Twitter-sphere: Graph-Based Approach", 2021 IEEE International Conference on Big Data (Big Data), pp.4136-4145.
- [11] K. Sutaria, D. Joshi, C. K. Bhensdadia and K. Khalpada 2015, "An Adaptive Approximation Algorithm for Community Detection in Social Network," 2015 IEEE International Conference on Computational Intelligence & Communication Technology, pp. 785-788.
- [12] T. K. Mackey, J. Kalyanam, T. Katsuki and G. Lanckriet 2017, "Twitter-based detection of illegal Online sale of prescription opioid", Amer. J. Public Health, vol. 107, no. 12, pp. 1910-1915.
- [13] H.-W. Meng, S. Kath, D. Li and Q. C. Nguyen 2017, "National substance use patterns on Twitter", PLoS ONE, vol. 12, no. 11.
- [14] J. Kalyanam and T. K. Mackey 2017, "A review of digital surveillance methods and approaches to combat prescription drug abuse", Current Addiction Rep., vol. 4, no. 4, pp. 397-409.
- [15] C. L. Hanson, B. Cannon, S. Burton and C. Giraud-Carrier 2013, "An exploration of social circles and prescription drug abuse through twitter", J. Med. Internet Res., vol. 15, no. 9, pp. e189.