

# Detection of Drug Abuse using Social media mining

Astika Kudali<sup>1</sup>, Siddhi Kunekar<sup>2</sup>, Prof. Renuka Nagpure<sup>3</sup>

<sup>1,2</sup>Dept. Information Technology, Atharva College of Engineering, Mumbai.

<sup>3</sup>Professor, Dept. Information Technology, Atharva College of Engineering, Mumbai.

\*\*\*

**Abstract** – A The graph of drug intake is seen to be growing nowadays. This is based upon surveys and other drug rackets which are caught in a not so effortless manner. People don't openly answer in such surveys to sensitive topics. Selection of people to survey also dictates the line of this resultant graph which is amongst other challenges. This information is vital to law enforcement agencies and for public welfare. In this paper, we uncover drug patterns using text and image mining. We extract dynamic data from social media platforms like Instagram to discover such wrong habitats. Initially, Instagram posts are mined using text analysis with the help of certain drug-related hashtags. The posts which are filtered are analysed to determine frequent drug-related slangs and linguistics, which further help in accessing more data. Similarly, other common trends and networking behaviors of the addicts are learnt using the obtained posts. User information and demographics is traced and using network analysis, other linked user-accounts are also inspected in similar manner.

**Key Words:** Social Media, Instagram, Data mining, Drug pattern detection, hashtags

## 1. INTRODUCTION

Drug activities are increasing a lot these days. Young children are involved in such illegal things. Behavior of children as well as adults is getting affected because of this. It is obviously not ideal for children to be involved in all this. To catch the drug dealers, it is difficult for the law enforcement agencies, as it takes a lot of planning and security to undergo the procedure. Another way of tracing these illicit activities is by conducting surveys. Now the conclusion of surveys depends upon many factors like representative sample undergoing the survey, also whether they are being completely honest in their answers, as people don't answer truthfully in concern for such sensitive subjects.

Data mining has a lot of applications. It aims to predict more accurate results, by which further diagnosis of a problem becomes easy. Data mining can be done for a lot of things. When data mining is done for a huge amount of data, it shifts towards more accurate results. Instagram as a social media platform is emerging as millions of photos are added into its data and large amount of people are joining the platform. Many important occasions of a person's life are recorded and stored in the database. More importantly, Instagram revolves around posting images about one's activities, and posts are usually accompanied by hashtags, single words that describe the activities/characteristics, which are what we use to fetch posts related to drug consumption.[1 ] Thus, Instagram acts as a dynamic source of real time data to give more

efficient results. This data can be of great use as a source to trace the illicit activities. Big data gives an opportunity to sort through wider range of people. Also, this can be done with limited sources and reduced time than other methods.

## 2. LITERATURE SURVEY

### 1.1 Fine-grained mining of Illicit Drug Use Patterns Using Social Multimedia data

This paper was proposed by Yiheng Zhou, Numair Sani, Jiebo Luo. The following paper proposes an idea of Instagram data mining for tracing prohibited use of drugs. This paper also uses image mining as well as text mining to implement the model. This paper has detailed information about the process used to carry forward the project. It is stated that it requires a lot of time to carry out traditional methods, thus it is suggested of a method which requires less of time and labor. Firstly, mine the data using hashtag-based algorithm and figure out time pattern of drug consumption. Various geographical trends were determined by location endpoint search, which in detail when used was giving much more accurate results. A basic development of this investigation is in mixed media information examination and particularly face picture examination. The ability to manage, analyse, summarize, visualize, and discover knowledge from the collected data in a timely manner and in a scalable fashion is a complex task. Storing this data without using it in any sense is simply a waste of storage space and time.

### 1.2 Detecting Topics and Locations on Instagram Photos

This paper uses Computer Vision tools and text mining techniques, and provides a strategy to identify similar content topics of pictures related to a specific hashtag on Instagram. Instagram is a social media platform that permits the users to share their status by transferring pictures with some caption, its location and a few hashtags.

According to this paper, the relevant content between posts was detected by following a procedure. Firstly, the web is scraped for all the posts related to a single hashtag. Also using Image recognition, relevant data is collected. Then using Visual description and User's description similar topics are recognized. Visual description meaning using Image analysis for detecting similar kind of posts and User's description meaning the caption of that post and the hashtags. Cosine method was computed to identify the similarity between the identified topics in the user posts description and the visual description in a value between 0 and 1; where 0 means there is no similarity and

1 means that both topics are exactly the same.[2] The results are similar to what was expected to be a result. For this reason, the methodology and procedure of this work can be valuable in the regions, for example, computerized promoting, statistical surveying, assessments of public sentiment, online networking, social investigations and different fields.

### 1.3 A methodology for mapping Instagram hashtag

It is expressed that it requires a great deal of time to perform customary strategies, subsequently it is proposed of a technique which requires less of time and work. The time for studying data through qualitative means is also critical to understanding how social media are used in nonstandard ways, identifying practices that might easily be missed through automated analyses.[3] Obviously, enormous data isn't synchronous with quantitative examination and there are increasingly significant bits of data to be studied from largescale testing and thus, we advocate for blended strategies, joining quantitative and subjective devices This paper enhances the methods that can be used for hashtag analysis on Instagram. Too often, Big Data enables the practice of apophenia: seeing patterns where none actually exist, simply because enormous quantities of data can offer connections that radiate in all directions.[3]

### 3. METHODOLOGY

**1.Scraping of data from social media:** Instagram is a social networking site where you can post pictures about anything with a caption and hashtags. The hashtags are basically to define what the picture is about. Mostly these hashtags help identify people to recognize which category the image belongs to. There are over 400 million people using Instagram, many of them post about their activities or achievements here. It is very difficult to trace a person in a real life. But on Instagram it is comparatively easy to identify the person. More importantly, Instagram revolves around posting images about one's activities, and posts are usually accompanied by hashtags, single words that describe the activities/characteristics, which are what we use to fetch posts related to what we want.[1] Also, on Instagram, it is easy to find connections between person to person, which just gives access to more data on Internet.

**2.Text Analysis:** Text Analysis is the computerized procedure of acquiring data from content. Content investigation is the mechanized procedure that permits machines to remove and order data from content, for example, tweets, messages, bolster tickets, item audits, overview reactions, and so forth. It is important to implement textual analysis so as to identify which tags are drug-related. In text analysis there are many technologies amongst which, only two methods are implemented which are used for our project.

Those are:

**Word Frequency:** In this it is checked, how many times the word occurs in a given text. It helps in finding out which tags are mostly used in those drug-related posts

and filter them out for further finding out about more tags to add in that list. **Collocation:** In some parts of text there might be multiple words that occur one after another. There are bigrams (Two words after one another) and trigrams (Three words after one another) which might coexist repeatedly which should also be counted. In our project, we see whether there are two or more hashtags that occur simultaneously in different posts to get the relativity in between them.

**3. Filtering posts and collecting the information:** There are many posts which are filtered according to different hashtags that are related to the drugs. We utilized an example where if a post has more than two hashtags which are in the rundown of hashtags that are mined from the Data mining and Text analysis, at that point those posts are considered as drug-related posts. From these posts, the hashtags which occur in more than six posts are considered as hashtags which are drug related. Then further these tags are added into the list of drug-related hashtags.

### 4. RESULT

Initially basic scrapping of data is done. Drug related images are showcased in the Web-Chrome.

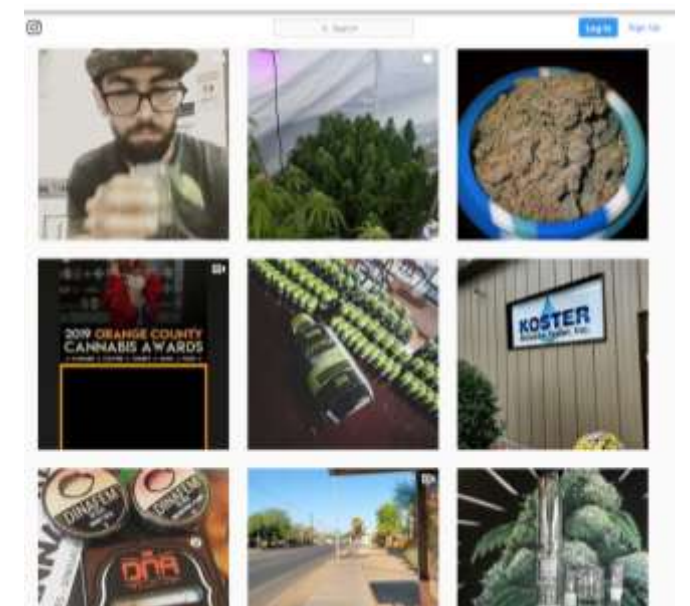
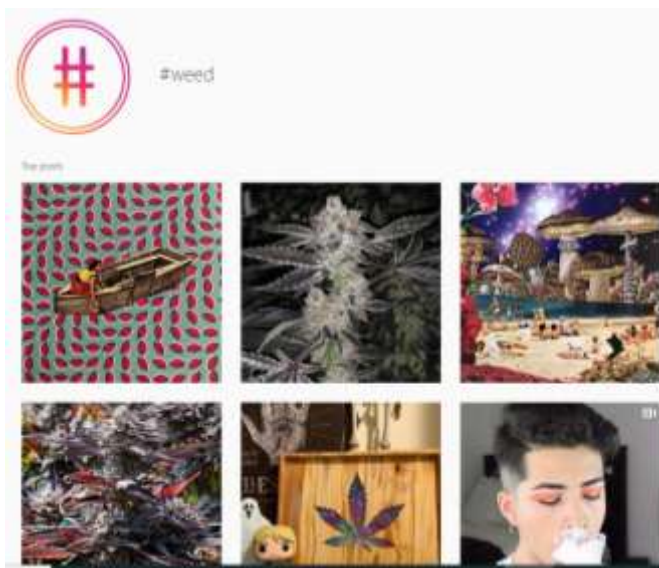


Fig-4.1: Instagram Drug related posts

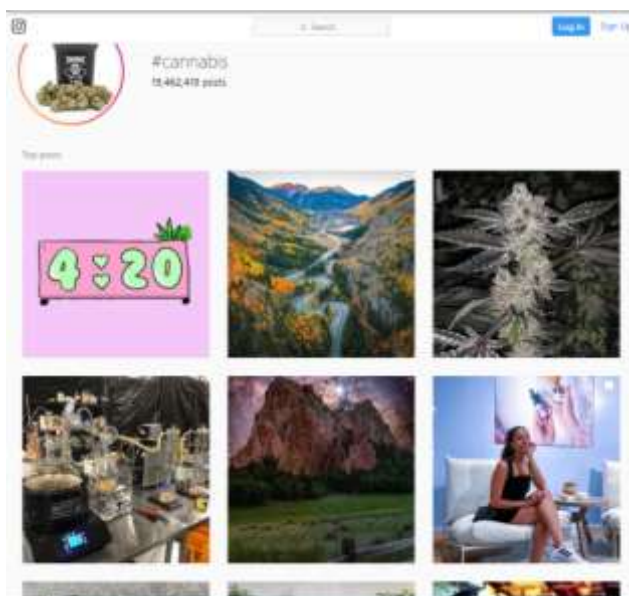
Then posts are filtered based on a particular hashtag

**Fig-4.2:** Based on hashtag weed

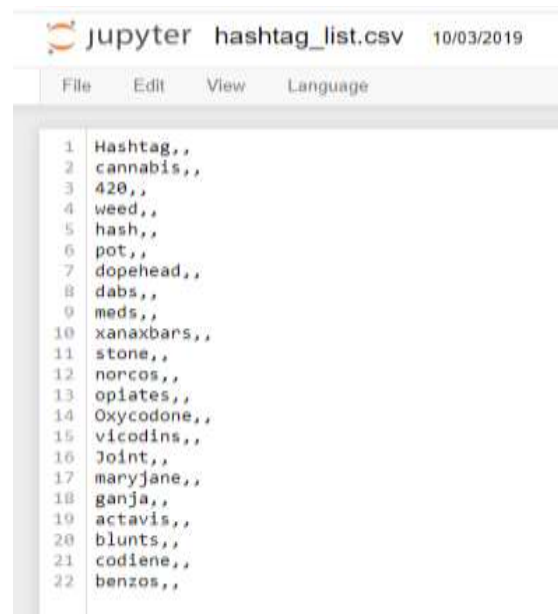
There are also other hashtags related to weed which gives other posts. In that way many hashtags are looked out for.

**Fig-4.4:** Instagram post with hashtags

There are other lingos which need to be found out through Instagram hashtags and other posts. The list is generated which consists of all drug related hashtags.

**Fig-4.3:** Based on hashtag cannabis

Some hashtags are when Instagram posts are filtered which are related to drugs. There are also hashtags mentioned in comments sometimes.

**Fig-4.5:**Hashtag list

## 5. CONCLUSION

Thus, we aim to develop a model that works on dynamic data from Instagram and detects the illegal drug related posts through the hashtags given in the description of the post. The model also runs through various other hashtags to learn the lingo of drug users. Similarly, other hashtags are added in a CSV file for further studying of their behavioral patterns. The user accounts which are found to be involved in such things are checked for such accounts which might be also involved in drug activities. Also, location-based analysis of posts is done for mining more of such illicit patterns.



## ACKNOWLEDGEMENT

We would like to thank our guide, Prof. Renuka Nagpure for helping us in giving an approach and guiding through the project. Without her support it wouldn't have been possible to implement this project in an efficient manner.

We are also grateful to our Principal Dr. S. P. Kallurkar and Prof. Deepali Maste (HOD, Information Technology Engineering) for giving us their insights and thoughts for the betterment of our project.

## REFERENCES

- [1] Yiheng Zhou, Numair Sani, Jiebo Luo, "Fine-grained Mining of Illicit Drug Use Patterns Using Social Multimedia Data from Instagram" IEEE, Nov. 2018.
- [2] Carlos Fiallos, Karina Jimenes, Stalin Figueroa "Detecting Topics and Locations on Instagram Photos" IEEE, Jan. 2016
- [3] Anju Mohandas, Bindhya Babu, Deepa Rajan S, Dr. L Padma Suresh, Rainu Boben "Survey on Mining Social Media Data for Understanding Drug Usage" IEEE, Jan. 2016
- [4] Anju Mohandas, Bindhya Babu, Deepa Rajan S, Dr. L Padma Suresh, Rainu Boben "Survey on Mining Social Media Data for Understanding Drug Usage" IEEE, Jan. 2016
- [5] Cody B., & Jennifer G. This is your Twitter on drugs. Any questions? 2015.
- [6] Munmun C., Michal G., Scott C., & Eric H. Predicting Depression via Social Media. 2013.