Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org

Volume 24, Issues 4, pp. 971–984, DOI 10.12694/scpe.v24i4.2471

A STATISTICAL ANALYSIS OF SENTIMENT OVER DIFFERENT SOCIAL PLATFORMS ON DRUG USAGE ACROSS HIGH, MIDDLE AND LOW-INCOME COUNTRIES

ANUREET CHHABRA[*], AKASH SHARMA[†], KUNAL CHHABRA[‡], AND AMIT CHHABRA[§]

Abstract. Social media serves as a platform for sharing information and connecting with others on various subjects, including healthcare and drugs. Analyzing drug sentiment from various social media like Twitter, reddit, quora etc are crucial for monitoring drug safety, identifying adverse reactions, and providing an early warning system for potential safety concerns, benefiting healthcare organizations and governments. This paper aims to study the opinions of people from all over the world by analyzing their messages, posts on social media. For this study 39,069 drug related message corpus were fetched making a comparison between the high income countries, and middle income and low income countries on the basis of drug consumption from 2021 to 2023. The dataset used for the study consisted of 41.63% text-corpus from high income countries out of which on an average from 2021 to 2023, 40.2% was found to have a positive sentiment. Whereas 34.65% of text-corpus are from middle-income countries out of which on an average 26.4% of was of a positive sentiment and 23.70% text-corpus are from low income countries with 23.6% having a positive sentiment. Furthermore, the primary factor for having such differences from people's sentiment on drug consumption from high to low income countries includes Cultural and Social Norms, Legalization, availability etc therefore, In high-income countries, drug use is more socially accepted than in other regions. This proposed study gives an insight into people's opinion on various drugs from different countries and regions. The results of this study attempted to understand how the public is responding to different types of information and to identify potential misinformation which can be used to formulate policies for existing and future drug prevention campaigns in order to improve public health and promote public education.

Key words: Sentiment Analysis, Drug consumption, drug overdose

1. Introduction. One of the finest places in the world right now for individuals to share their opinions on a given topic are Twitter, Reddit, Quora etc [1]. Social media platforms are widely used by individuals, businesses, and organizations to share news, opinions, and updates with their followers, and is a popular platform for real-time news and discussions on a wide range of topics. Also these platforms provide access to its API (Application Programming Interface), which is a set of tools and protocols that developers can use to access and interact with data and functionality [2]. Due to APIs of these platforms, they have become a significant source of real-time data that can be utilized in studies to analyze sentiments of the people towards social and political issues. Currently, an increased consumption of drugs and alcohol by youth is a growing concern among parents, educators, and healthcare professionals. Drugs and alcohol can have severe negative effects on individuals and society as a whole. They are highly addictive and can lead to physical and psychological dependence [3]. The increased consumption of alcohol by teenagers and youth can have serious negative consequences, including increased risk of accidents and injuries, impaired judgment and decision-making, and increased risk of alcohol-related health problems [4]. The worldwide deaths due to drug overdose are exponentially increasing as shown in figure 1.1 and deaths due consumption of different drugs in the USA is shown in figure 1.2. The rate of smoking and tobacco is very high as per the latest surveys as shown in figure 1.3 and figure 1.4 respectively. Excessive drug consumption can lead to addiction, legal problems and social problems. Narcotics drug abuse can also lead to overdose and death. In addition to the harm caused to individuals, narcotics drug abuse can also have significant social and economic costs. It can lead to increased crime and reduced productivity, and can put a strain on healthcare and criminal justice systems [5].

Sentiment analysis on drug-related messages, posts etc can be used to develop targeted interventions and prevention strategies, such as educational campaigns or treatment programs. For example, if sentiment analysis

[*]Chandigarh College Of Engineering and Technology, Panjab University. (anureet0409@gmail.com)
[†]Chandigarh College Of Engineering and Technology, Panjab University (sharmaaakash080@gmail.com).
[‡]Vellore Institute of Technology (kunalch7677@gmail.com).
[§]Chandigarh College Of Engineering and Technology, Panjab University. (amitchhabra@ccet.ac.in).

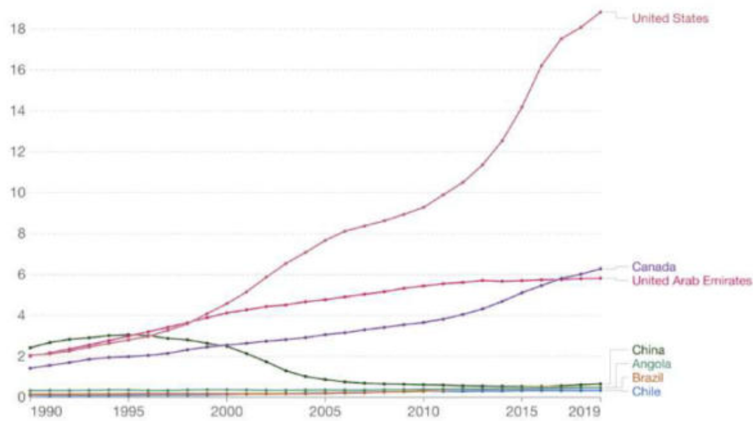

Fig. 1.1: Country wise Drug Use disorder death rate (Per 100,000) on Y›-axis, Year (1980 to 2019) on X-axis[6].

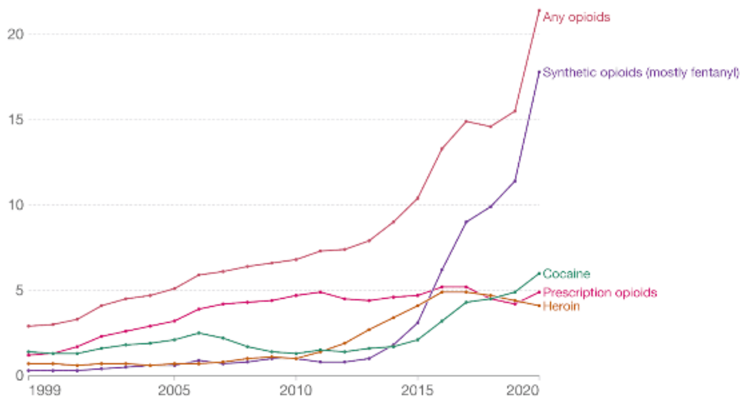

Fig. 1.2: Deaths due to drug overdose in the USA (Per 100,000) on Y-axis from the year 1999 - 2020 (X-axis) [6].

reveals an increase in the use of a particular drug, targeted educational campaigns could be developed to raise awareness of the risks associated with that drug and discourage its use. In addition, sentiment analysis can be used to monitor the effectiveness of prevention strategies and treatment programs over time. Moreover, through the analysis of drug-related messages, researchers can identify various concerns such as driving under the influence and underage drinking, substance abuse, and drug intoxication. Which can be valuable in developing effective public health policies [7].

The aim of this study is to present a detailed sentiment analysis of drug-related posts, messages at a country and also at a global level. The methodology involved in this paper utilizes several libraries, including Snscrape, Tweepy, NLTK, TextBlob, and API's of Twitter, Quora and reddit to generate the desired result. NLTK is an open-source library which provides a wide range of tools and resources for natural language processing, including tools for sentiment analysis. The further sections are divided as follows. Section 2 contains the review of the existing work on drug related sentiment analysis. Section 3 discusses the societal impact of this study. Section 4 contains the methodology followed to collect the data and perform sentiment analysis on it. Section 5 contains the results and analysis for both global and country wise segmented data and discussion about the results obtained are present in section 6. Section 7 contains the limitations and the future directions of this study.

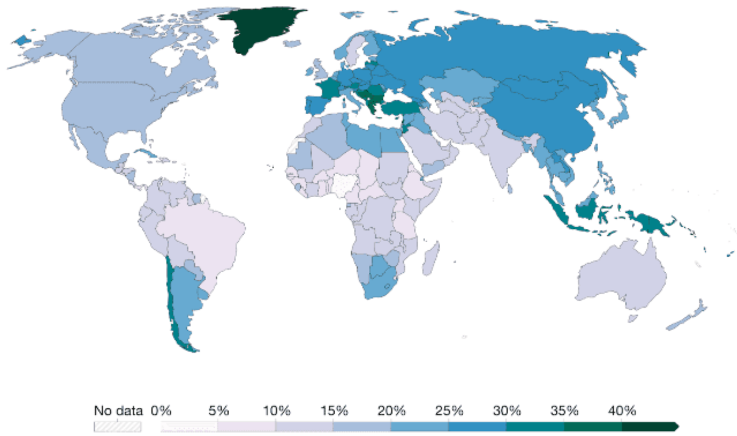

Fig. 1.3: Country-wise Prevalence of Global daily smoking populations (people of age 15 years and older) for the year 2021 [6].

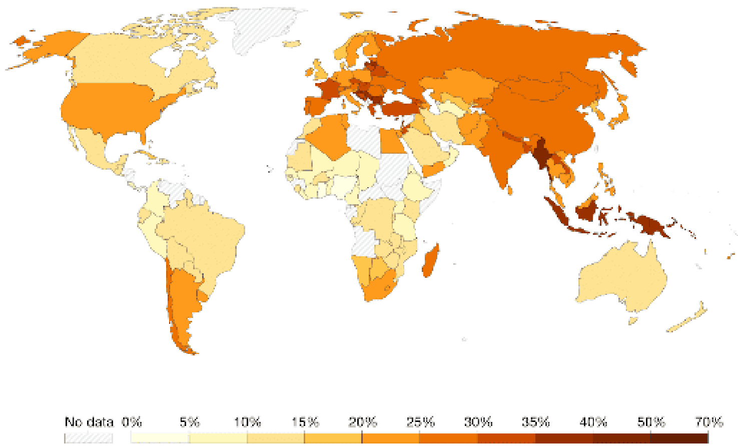

Fig. 1.4: Global use of tobacco (age 15 years and older) for the year 2020 [6].

2. Literature Review. The research [8] explores the use of sentiment analysis of tweets to extract information about adverse drug events (ADEs). The study uses a combination of natural language processing and machine learning techniques to analyze tweets and to identify ADEs related to specific drugs. The study found that sentiment analysis can be an efficient and effective way to extract information about ADEs from tweets, and that the proposed method was able to identify ADEs with high precision and recall. The research shows that sentiment analysis of tweets can be a useful tool for monitoring ADEs and for identifying potential safety issues related to specific drugs. In [9] the study looks at the impact of sentiment analysis on extracting adverse drug reactions (ADRs) from tweets and forum posts. The study found that sentiment analysis can be used to identify specific ADRs that are not reported in traditional sources such as clinical trials. The research [10] looks at the use of sentiment analysis on tweets to understand public opinion of the application of drugs in different countries but no categorisation has been made. The study uses textblob for identifying the sentiment of tweets in different languages and different countries.

In [11] the use of sentiment analysis and transfer learning to improve the classification of adverse drug reactions (ADRs) has been used. The study found that the use of transfer learning from a pre-trained model can be effective in handling the class imbalance present in the ADR data. The study [12] explores the use of sentiment analysis on Twitter data to extract information about adverse drug reactions (ADRs). The study found that sentiment analysis can be an effective way to extract ADRs from tweets and that the proposed method was able to identify ADRs with high precision and recall. The study also found that the use of sentiment analysis can be useful for monitoring ADRs and identifying potential safety issues related to specific drugs. The study concludes that sentiment analysis of tweets can be a useful tool for discovering ADRs and for identifying potential safety issues related to specific drugs. Study [13] looks at public opinion of the application of drugs in different countries by analyzing tweets. The study uses VADER Sentiment Analysis, a pre-trained lexicon-based model, to classify tweets according to sentiment (positive, negative, or neutral) and to determine the viewpoint of the tweets. The study found that the sentiment of tweets varies depending on the country and the drug being discussed, and that VADER can effectively classify the sentiment of tweets in different languages and different countries. The study also found that the use of VADER can provide a quick and efficient way to understand public opinion of the application of drugs in different countries.

In study [14] presents a comprehensive review and benchmark evaluation of the state-of-the-art techniques for sentiment analysis of tweets is given. The study provides a benchmark evaluation of the performance of various techniques on a dataset of tweets. The study found that the most effective techniques for sentiment analysis of tweets are deep learning approaches such as convolutional neural networks and recurrent neural networks. The study also found that these techniques performed better than traditional machine learning approaches such as support vector machines and decision trees. The study [15] is a survey and comparative study of different semi-supervised learning techniques for sentiment analysis of tweets. The aim of the study is to evaluate the performance of these techniques and to compare their performance. The study found that the semi-supervised learning techniques performed well on the sentiment analysis task, with the best performance achieved by the proposed method, which combined a combination of feature extraction, feature selection, and classification.

The study [16] presents a Sentiment analysis on gabapentinoid (pregabalin and gabapen- tin) has been performed. Some keywords taken are: Pregabalin, Lyrica, gabapentin, Neurontin. From 8 March to 7 May 2021. They have extracted 2931 pregabalin-related tweets out of which 30% are found to be having a positive sentiment, 21% are found to be negative and 2736 gabapentin-related tweets out of which 22.7 are found positive, 9.1% are found to be negative. Sentiment analysis on drugs used for therapy of COVID-19 using VADER is performed in paper [17] and a dictionary based approach. Keywords taken for in study are Hydroxychloroquine etc. on which 12.8%, 13.33% and 13.5% of tweets are found to be of positive sentiment in India, USA, China. Out of many countries African countries were found to be having maximum number of positive tweets with 18.3% In Study [18] authors have improved precision and recall of sentiment analysis on data from DailyStrength forum and tweets with a total messages of 8051. Keywords taken for drug related sentiment analysis are Trazodone, Cymbalta, Quetiapine, Seroquel. Proposed approaches by the authors have increased F-measure to 69.16% and 80.14% from twitter and DailyStrength respectively. The study [19] presents a total 267,215 tweets from Nov-2014 to Feb-2015 have been extracted using 250 drug-related tweets keywords such as adderall, xanax, Aspirin etc. Although the study has not done sentiment analysis but rather the authors focuses on ADE in which they found the maximum tetra-gram language model scores to be 1 and minimum score to be 0.21.

The above mentioned literature survey does not provide a comprehensive analysis based on income division and population of countries which indirectly affect education in the country. Furthermore, the available studies don't provide sentiments of people on drug related keywords that are taken in this study from social media platforms such as Twitter, Reddit, Quora etc. To address the limitations in the existing literature, our research aims to conduct a comparative sentiment analysis of drug-related message corpus like posts across low, middle, and high-income countries. By analyzing sentiments from diverse income-level nations, our work seeks to provide a more comprehensive and representative view of public opinion on drug usage, efficacy, and societal acceptance. Additionally the study also provides a global sentiment analysis on a wide range drug related keywords like “heroin”, “cocaine”, “nicotine”, “smoking” etc

3. Societal Impact of the Study. Drug sentiment analysis on social media messages, tweets, posts can help society in several ways. By analyzing messages related to drugs, researchers in the current study provides insight into public opinion on various drugs and the potential side effects of these drugs [20]. This

Table 4.1: Sample Dataset generated using APIs of Twitter, Reddit, Quora

Date	Content	Location	Sentiment
2022-12-08 05:32:24+00:00	I can't wait to go back to university and commit Zina, listen to music, do drugs, go to parties, etc. and most importantly freemix	USA	Positive
2022-12-03 14:40:05+00:00	Mumbai: Man stabs friend to death during quarrel over paying for alcohol, held	India	Negative
2022-12-08 14:24:13+00:00	This should have never happened. Seems like the attacker has a problem with alcohol	USA	Negative
2022-04-10 03:19:06+00:00	Best Alcohol Drink in Euphoria??	USA	Positive
2022-12-08 14:34:43+00:00	It's Thursday, which means it's almost Friday, which means that the alcohol delivery is set to arrive any minute. I'm hyped	USA	Positive

information can be used to inform public health policy and to identify potential safety concerns related to drugs. This study can be used by pharmaceutical companies to monitor the public's perception of their drugs and to identify potential issues related to drug development and marketing [21]. The Sentiment analysis in this study aims to provide most common misconceptions and misunderstandings about drugs, which can then be addressed through public education campaigns.

Furthermore, Through this study early signs of negative public sentiment towards a drug can be identified, which can be used as an early warning system to allow pharmaceutical companies and regulatory agencies to take action before a safety concern becomes a major public health issue [22,23]. Sentiment analysis can be used to monitor the impact of drug-related news and events on public sentiment. This can be used to understand how the public is responding to different types of information and to identify potential misinformation [24]. Sentiment analysis on drugs related messages can provide valuable information to help improve public health, inform drug development and marketing, and promote public education. Misinformation about drugs on social media can be identified using sentiment analysis, and this information can be used to inform public education campaigns [25].

4. Methodology.

4.1. Data Collection. The dataset consists of 39,069 message corpus extracted or scraped from reddit, twitter and quora using their APIs. The dataset's properties include date-time, content, location and sentiment. The dataset comprised of 40% tweets from twitter and 60% of the data from reddit and quora. The dataset has been categorized to represent negative sentiment, neutral sentiment, and positive emotion which is due to the combined result of cultural, social, accessibility, availability and other multiple factors as further discussed in section 6. Negative sentiment within the dataset reflects the disapproval and condemnation of drugs by individuals, while positive emotion corresponds to instances where people express happiness and approval of drug consumption, often seen in celebratory contexts and similar situations. The neutral sentiment category, on the other hand, captures instances where drug-related sentiments neither lean towards negativity or positivity. A sample of the records has been shown in table 4.1. The messages are sourced from India, Mexico, USA, Russia, Australia, Bangladesh, Sri Lanka which have been segmented into low, middle and high income and population wise as shown in Table 4.2 to facilitate a comprehensive analysis and obtain a generalized demographic overview.

4.2. Data Preprocessing. Initially, messages such as posts related to drugs and alcohol are gathered from twitter, reddit and quora. The gathered messages corpus are cleaned, and prepared. We utilized SNSCRAPE and TWEEPY, Reddit and Quora's API to search for text concerning the drug such as Morphine, Ganja, Alcohol etc. The option "near" is set to the country on which we are doing analysis to get a message with geolocation coordinates inside or close to that country. The language option "Lang" to "en" to get texts/messages in English. Apart from the text collected from social media platforms, the data associated with each text is also collected—such as the text's location, links to pertinent media, or user mentions—represents a sizable quantity of information. The overall methodology has been shown in figure 4.1.

Table 4.2: Countries selected for the case study based on Population and Income as a factor.

Country	**Population Level**	**Income**
India	High	Middle
Mexico	Middle	Middle
USA	High	High
Russia	High	High
Australia	Low	High
Bangladesh	High	Low
Sri Lanka	Low	Low

Data cleaning is necessary prior to sentiment analysis. Techniques for data preparation have evolved across a wide range. The Python Natural Language Toolkit is one of several tools for minimizing text format complexity [26]. Retrieve raw data, operate with HTML elements, and parse text using regular expressions . It also does text stemming, tokenization, and text similarity calculations [27]. Exclude terms and phrases that give no information as part of our data preprocessing, such as (is, the, a, etc). NLTK is used for data preparation in our study. Text similarity as a preprocessing approach minimizes the amount of data collected. The value of the data set is increased by carefully removing duplicate texts with a lot of similarities using a text similarity measure. To preprocess the data links, emojis and punctuation are removed from the text. Then the preprocessed text is converted to ASCII values to pass it to the VADER tool for sentiment analysis.

4.3. Tokenization of Text. Tokenization is a rapid method for converting raw input into a usable data string. Even though tokenization is most typically utilized in the building of NFTs and cybersecurity, it is still an important stage in the NLP process [28]. Tokenization is a NLP approach that breaks down phrases and paragraphs into more manageable, language-assignable bits. The first steps in the NLP process involve obtaining information (a phrase) and breaking it down into digestible chunks [29].

In this study we used VADER (Valence Aware Dictionary for Sentiment Reasoning) to do sentiment analysis and categorize retrieved texts as positive, negative, or neutral as shown in algorithm 1. VADER takes into account both emotional polarity (positive/negative/neutral) and intensity (strong). Using the wisdom of the crowd and human raters to validate the text sentiment analysis, VADER text sentiment analysis was introduced in 2014 [30]. Based on the amount of exclamation points and question marks at the conclusion, VADER sentiment analysis increases the sentiment score of the phrase. The VADER algorithm accounts for word capitalization by increasing or decreasing its emotion score by 0.733 based on the term's positive or negative meaning.

4.4. Generation of Word Clouds. Simple APIs are provided to enable common NLP tasks, including part-of-speech tagging, noun phrase extraction, sentiment analysis, classifying, and translating [31]. It provides straightforward APIs for several NLP applications, including sentiment analysis, noun phrase extraction, parts-of-speech tagging, translation, and classification. To determine the polarity of the messages, TextBlob has been used. Tools for sentiment analysis, pattern recognition, and natural language processing are available in TextBlob. 2,918 terms make up TextBlob's emotion vocabulary [32]. When using TextBlob's sentiment analysis tool, it is possible to determine whether a text contains true information or someone's opinion by looking at its polarity and subjectivity scores.

5. Result and Analysis. In this case study, the analysis has been segmented into three distinct categories based on the countries' income and population levels: high, middle, and low. In order to present a comprehensive understanding of how drug sentiment varies across different economies and countries over the years. This drug sentiment analysis is provided to gain valuable insights into the prevailing attitudes and perceptions towards drugs in each group and how the change over time. Moreover, to present an overview, This study has also been extended to global study that encapsulates the trends and patterns in drug sentiment across the entire spectrum of countries. As per the word clouds generated, the most frequent words obtained for each sentiment has been summarized in table 5.1.

ALGORITHM 1 Texts/messages Scraping from APIs and Sentiment Analysis using VADER

Input: Extracted Drug Related message using APIs

Output: Predicted Emotion of the messages as either positive/negative or neutral

Begin:

```
Import Libraries
Message_Corpus ← Initialize Empty List
While i < Number of Messages to be scraped
      em= Extracted messages using APIs with keywords "Ganja" , "alcohol","drugs"...
      Message_Corpus .append(em);
End While
Remove redundant messages in the dataset
Preprocess the messages by removing links and emojis
message_df = Convert Message_Corpus to Dataframe with Attributes "message",
"date-time", "Location","Sentiment"
For i in message_df
          IF Sentiment_polarity_score(message_df["message"][i]) >= 0.05 then
                    message_df["Sentiment"][i] = Positive
          Else
                    IF Sentiment Polarity Score (message_df ["message"][i]) <= -0.05 then
                              message_df ["Sentiment"][i] = Negative
                    ELSE
                              message_df ["Sentiment"][i] = Neutral
                    END IF
          END IF
END FOR
```

Table 5.1: Most frequent words in each sentiment for Global analysis

Negative	Positive	Neutral
Suicide	Recreational	Legalization
Depress	Nice	Food
Addiction	Sweet	Morning
Abuse	Enjoy	Medication
Dangerous	Celebrate	Treatment
Overdose	Appreciate	Dosage
Deadly	Party	Government

5.1. Global Analysis. For the global data analysis of drug related text sentiment analysis in this paper keywords such as "alcohol", "Ganja", "nicotine" have been used in the APIs. The word cloud of each keyword has been visualized from figure 5.1 to figure 5.4. The opioid crisis has been a significant public health issue in many countries, particularly the United States along with other countries as seen in figure 1.3. Opioids, including prescription painkillers and illegal drugs like heroin and fentanyl, have led to a surge in overdose deaths and addiction cases. Although the smoking rates have been declining in many countries, especially in developed nations, the numbers are still significant. This decline is due to public health campaigns, increased awareness of the health risks associated with smoking, and the implementation of tobacco control policies such as higher taxes, smoking bans, and graphic warning labels which can be enhanced further. Excess of Tobacco consumption is a major challenge in many countries as shown in figure 1.4. Tobacco use remains a leading cause of preventable deaths worldwide. It is estimated that millions of people die each year due to tobacco-related

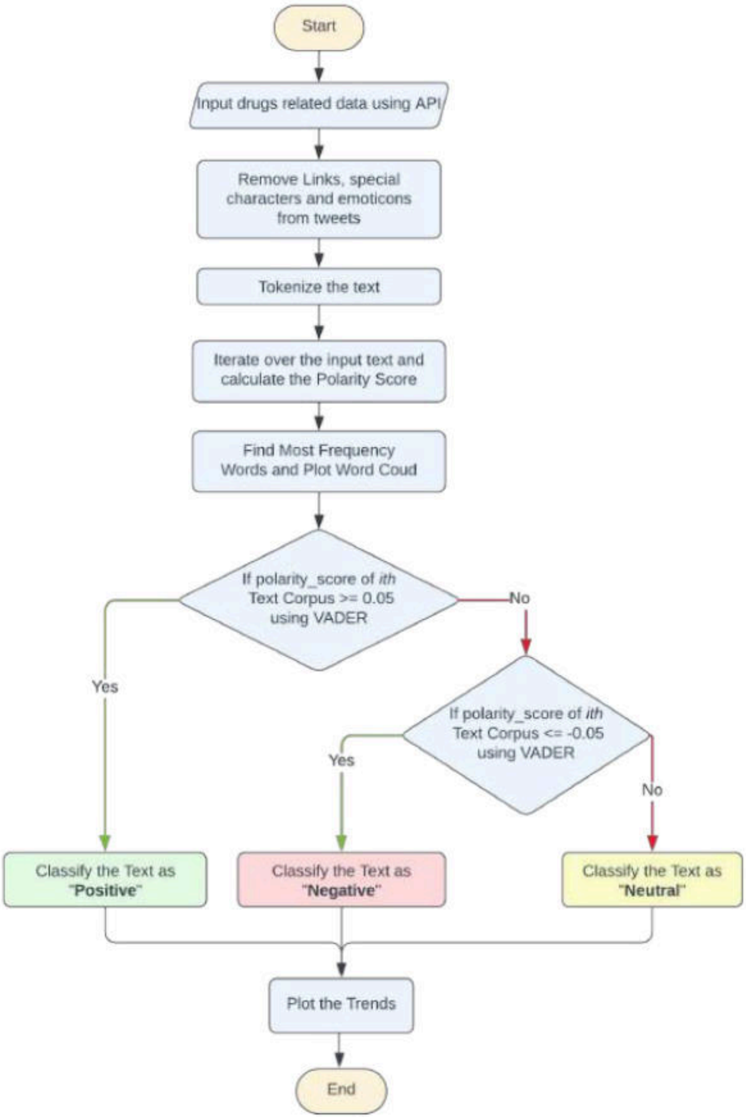

Fig. 4.1: Methodology followed for drug related posts or text Sentiment Analysis.

Fig. 5.1: Word-cloud of keyword “ganja” (Left) and “alcohol” (Right) on the global dataset.

Fig. 5.2: Word-cloud of keyword "nicotine" (Left) and "cigarette" (Right) on the global dataset

Fig. 5.3: Word-cloud of keyword "smoking" (Left) and "Heroin" (Right) on the global dataset

illnesses, including lung cancer, heart disease, respiratory problems, and other health issues. Furthermore in table 6.1 global drug sentiments has been shown for each keyword and the word-wise sentiments in the texts is presented in table 5.1. Furthermore, the analysis for the data of global messages or posts are present as follows in table 5.2. Various keywords are taken such as ganja, cocaine, smoking nicotine etc.

5.2. Income as a factor. High-income countries, also known as developed countries, are nations with a high gross national income (GNI) per capita. These countries typically have a well-developed infrastructure, a high standard of living, and a strong economy. Some examples of high-income countries include: United States, Russia, Australia etc. Low-income countries, also known as developing countries, are nations with a low gross national income (GNI) per capita [33]. These countries typically have a less developed infrastructure, lower standard of living, and a weaker economy. Some examples of low-income countries include: Bangladesh, Sri Lanka etc.

In high-income countries, drug consumption is higher than in lower and middle-income countries. The use of opioids, such as heroin, is a major concern in these countries. According to the World Health Organization (WHO), opioid use disorders affect an estimated 14 million people globally, with the highest prevalence in North America and Western Europe. The use of other types of drugs such as cocaine and amphetamines is also commonly consumed in high-income countries [34]. The dataset generated for the study consisted of 41.63% (16,268) text from high income countries (USA, Russia, Australia) whereas the rest 34.65% (13,538) text are from middle-income countries like (India, Mexico) and 23.70% (9,263) text are from low income countries (Bangladesh, Sri Lanka).

In the generated dataset we observed that the average as follows: 35.60% of the extracted text from high income countries (USA, Australia, Russia) were classified as neutral and 40.02% as positive from the year 2021-2023 which is a matter of concern. As seen in the dataset, in high income countries, drugs and alcohol is a major part in every celebration and in their daily lives. Whereas In middle-income countries, drug and alcohol consumption tends to be lower than in high-income countries but higher than in low-income countries. The

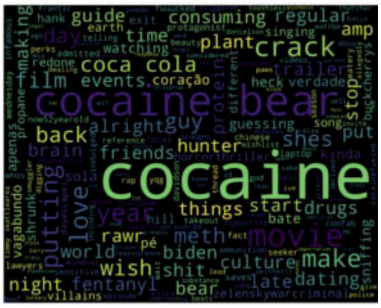

Fig. 5.4: Word-cloud of keyword “cocaine” in the global dataset

Table 5.2: Most frequent words in each sentiment for Global analysis

Sr. no.	Keyword	Positive Tweets	Negative Tweets	Neutral Tweets
1	Ganja	30.49 %	9.88%	59.63%
2.	Alcohol	33.41%	14.86%	51.73%
3.	Nicotine	15.0%	15.0%	70.0%
4.	Cigarette	29.83%	23.91%	46.26%
5.	Smoking	29.837%	16.599%	53.564%
6.	Heroin	17.018%	23.376%	59.606%
7.	cocaine	20.158%	18.346%	61.496%

use of opioids and cannabis is relatively common in some countries. The non-medical use of prescription drugs is also a concern. Through detailed analysis we have observed that on an average of the three years, 44.10% of the text were neutral with 26.40% of text turned out to be negative in middle income countries . In Low income countries the average sentiments are summarized as follows, 35% of text are Positive and 23.6% of text are Neutral. The pie charts summarizing and comparing the results of low, middle and high income countries average from 2021 to 2023 are shown in figure 5.5.

In low-income countries, drug consumption tends to be lower than in high-income and middle-income countries. However, the production and trafficking of narcotics drugs is a major concern in many of these countries like Afghanistan. Many low-income countries are major producers of opium and coca, the raw materials used to make heroin and cocaine, respectively. The illegal trade in these drugs generates huge profits, fueling corruption and destabilizing governments [35]. Drug consumption in low-income countries is also a concern, especially in urban areas.

6. Discussion about the Results. The drug-related sentiment of people from the year 2021 to 2023 from countries with income (high, middle, low) as a factor has been summarized in table 6.1 and year wise analysis has been shown in figures 6.1 - 6.3. Which shows that the although consuming drug creates addiction and is morally negative but still there is a lot of positive and neutral sentiment of people towards it. In this study various such factors are studied to understand the sentiment of people towards drugs. The primary factor that impacts the drug-use of the people in a nation includes Socio-economic, Cultural and Social Norms, Legalization and Decriminalization etc as discussed below in this section.

Due to socio-economic factors in populated and more developed nations, income disparities and social inequalities can lead to marginalization, where drug use might be more prevalent as a coping mechanism. Through the results of this study it has been found that in high income and middle income countries the percentage of positive and neutral sentiment is increasing throughout the years whereas in low income countries positive and neutral sentiment is decreasing. The projected values till 2023 are discussed as follows. In high income countries positive sentiment increased from 38.52% to 41.46% and negative sentiment 26.51% to 21.97% from 2021 to 2023. In middle income countries, the positive sentiments increased from 25.38% to 27.27%

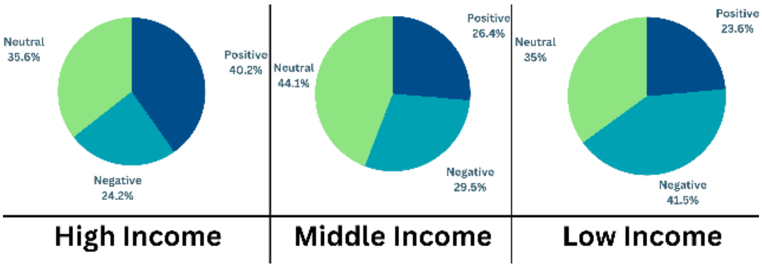

Fig. 5.5: Pie charts for results of of High, Middle and Low income countries average from year 2021 to 2023

Table 6.1: Drug related text sentiments of countries with income as a factor from year 2021 to 2023

Classification	Year	Positive Tweets	Negative Tweets	Neutral Tweets	Total Tweets
High Income	2021	38.52%	26.51%	34.97%	16,268
	2022	40.65%	24.07%	35.28%	
	2023	41.46%	21.97%	36.57%	
Middle Income	2021	25.38%	30.8%	43.74%	13,538
	2022	26.40%	29.48%	44.12%	
	2023	27.27%	28.15%	44.58%	
Low Income	2021	24.28%	40.33%	35.39%	9,263
	2022	23.91%	40.65%	35.24%	
	2023	22.37%	43.35%	34.28%	

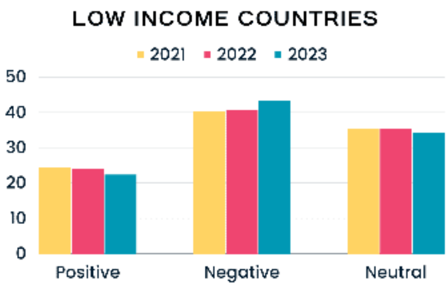

Fig. 6.1: Year wise sentiment analysis of Low Income countries

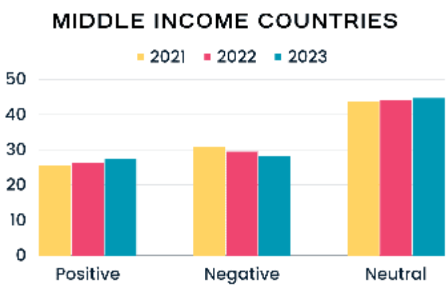

Fig. 6.2: Year wise sentiment analysis of Middle Income countries

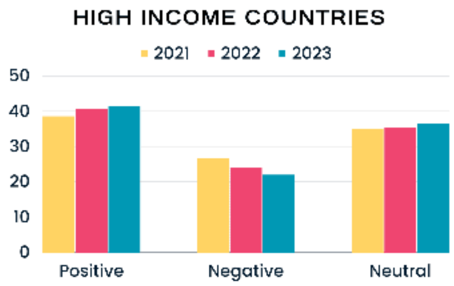

Fig. 6.3: Year wise sentiment analysis of High Income countries

and the negative sentiment decreased from 30.8% to 28.15% from the year 2021 to 2023. On the other hand, in smaller and medium-sized countries, the distribution of wealth and access to resources is more balanced, potentially resulting in lower overall drug consumption rates. Furthermore, Different societies and cultures have varying attitudes and norms towards drug use. In high-income countries, drug use is getting more socially acceptable with time or perceived differently than in other regions which is perhaps the reason for increasing positive sentiment on drug consumption. In low income countries positive sentiment decreased from 24.28% to 22.37% and the negative sentiment is increasing from 40.33% to 43.35% from the year 2021 to 2023. This trend signifies that people are discouraging drug usage and not accepting the western culture or norms from high and middle income countries. High-educated countries have deeply ingrained cultural norms that discourage drug use, leading to lower consumption rates regardless of population size. Conversely, in countries where drug use is more normalized, higher consumption rates may be observed. Understanding cultural factors and social norms is crucial for designing effective drug prevention and harm reduction strategies.

In some high-income countries, the people consume more drugs as compared to other countries due to legalized or decriminalization of specific drugs. This can lead to a shift in public perception and a more open discussion about drug use, which is reflected in positive sentiments of the input text message. Moreover, The availability and accessibility of drug treatment and support services can play a pivotal role in mitigating the impact of drug consumption. Smaller and under-developing countries face challenges in establishing comprehensive treatment programs due to limited resources, whereas larger countries may struggle to reach all affected individuals effectively. Understanding the relationship between population size, the prevalence of treatment services, and their effectiveness in reducing drug abuse is crucial for improving public health interventions. Smaller countries are more vulnerable to drug trafficking due to their limited resources for border security and law enforcement . The influx of illicit drugs contributes to higher rates of drug consumption and related problems. On the other hand, larger countries have more extensive border control measures but can still face significant challenges in combating drug trafficking.

7. Limitation and Future Scope. The current research has certain limitations that should be acknowledged. Firstly, the retrieved messages might not be exhaustive, as some drug-related data could have been removed by the respective social media platforms in accordance with their community guidelines. Moreover, the analysis relies on Text Mining techniques, inadvertently disregarding special characters, emojis, and images that may carry significant contextual information or emotional expressions related to drug use. However, it's not always possible to accurately interpret the sentiment of a message, and there may be biases in the data that affect the analysis. Additionally, sentiment analysis can only provide a snapshot of public perception at a particular point in time, and the sentiment of social media data may change over time. Despite these limitations, drug sentiment analysis can be a valuable tool for healthcare professionals and researchers who are looking to stay informed about public perception of drugs and develop effective strategies for drug development, marketing, and safety.

8. Conclusion. Drug sentiment analysis on data collected from various social media like twitter, reddit and quora provides valuable insights into public drug perception, monitoring safety, and enabling effective strategies for government and non-profit organizations to take necessary actions. This study examined drug-related sentiments of people over various social media platforms in different countries. It was found that in high income countries 40.2%, and in low income countries 23.6% of the average input text corpus from the year 2021-2023 were found to be positive, Which shows that people on the internet are getting more inclined towards a positive sentiment for drugs and alcohol yearly in high and middle income countries. Whereas a decreasing trend is seen in low income countries which signifies that people are not promoting drug usage. Drugs are a major part of celebrations and parties in the current world. Various factors, including socio-economic disparities, cultural norms, drug legalization, and treatment availability, influence drug consumption patterns. Smaller countries may face challenges in establishing comprehensive treatment programs and combating drug trafficking, whereas some developed countries have legalized various drugs and alcohol making it accessible to common people. Analyzing these trends is essential in designing targeted drug prevention and harm reduction strategies for different regions, promoting public health, and ensuring the well-being of communities worldwide.

REFERENCES

[1] Singh, D. K. S., Nithya, N., Rahunathan, L., Sanghavi, P., Vaghela, R. S., Manoharan, P., ... & Tunze, G. B. (2022). Social network analysis for precise friend suggestions for twitter by associating multiple networks using ml. International Journal of Information Technology and Web Engineering (IJITWE), 17(1), 1-11.

[2] Biswas, S., & Ghosh, S. (2022). Drug Usage Analysis by VADER Sentiment Analysis on Leading Countries. Mapana Journal of Sciences, 21(3).

[3] Saraswat, M., Dubey, A. K., & Gupta, R. Sentiment Analysis of Drugs related Post Impacting Society Healthcare.

[4] Abbasi, N., Vahdat-Nejad, H., Mansoor, W., & Salmani, F. (2022, December). Extracting Drug-Related Tweets in COVID-19 Pandemic. In 2022 5th International Conference on Signal Processing and Information Security (ICSPIS) (pp. 103-106). IEEE.

[5] Singh, M., Dhillon, H. K., Ichhpujani, P., Iyengar, S., & Kaur, R. (2022). Twitter sentiment analysis for COVID-19 associated mucormycosis. Indian Journal of Ophthalmology, 70(5), 1773-1779.

[6] Hannah Ritchie, Pablo Arriagada and Max Roser (2022) - "Opioids, cocaine, cannabis and other illicit drugs". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/illicit-drug-use' [Online Resource]

[7] Panda, B., Panigrahi, C. R., & Pati, B. (2022). Exploratory data analysis and sentiment analysis of drug reviews. Computación y Sistemas, 26(3), 1191-1199.

[8] Peng, Y., Moh, M., & Moh, T. S. (2016, August). Efficient adverse drug event extraction using Twitter sentiment analysis. In 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (pp. 1011-1018). IEEE.

[9] Garg, S. (2021, January). Drug recommendation system based on sentiment analysis of drug reviews using machine learning. In 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence) (pp. 175-181). IEEE.

[10] Vijayaraghavan, S., & Basu, D. (2020). Sentiment analysis in drug reviews using supervised machine learning algorithms. arXiv preprint arXiv:2003.11643.

[11] Alhuzali, H., & Ananiadou, S. (2019, August). Improving classification of adverse drug reactions through using sentiment analysis and transfer learning. In Proceedings of the 18th BioNLP Workshop and Shared Task (pp. 339-347).

[12] Bose, D., Aithal, P. S., & Roy, S. (2021). Survey of Twitter Viewpoint on Application of Drugs by VADER Sentiment Analysis among Distinct Countries. International Journal of Management, Technology, and Social Sciences (IJMTS), 6(1), 110-127.

[13] Moh, M., Moh, T. S., Peng, Y., & Wu, L. (2017). On adverse drug event extractions using twitter sentiment analysis. Network Modeling Analysis in Health Informatics and Bioinformatics, 6(1), 1-12.

[14] Zimbra, D., Abbasi, A., Zeng, D., & Chen, H. (2018). The state-of-the-art in Twitter sentiment analysis: A review and benchmark evaluation. ACM Transactions on Management Information Systems (TMIS), 9(2), 1-29.

[15] Mathieson S, O'Keeffe M, Traeger AC, Ferreira GE, Abdel Shaheed C. Content and sentiment analysis of gabapentinoid-related tweets: An infodemiology study. Drug Alcohol Rev. 2022 Dec 20. doi: 10.1111/dar.13590. Epub ahead of print. PMID: 36539307.

[16] Bose, Rajesh & Aithal, Sreeramana & Roy, Sandip. (2021). Survey of Twitter Viewpoint on Application of Drugs by VADER Sentiment Analysis among Distinct Countries. International Journal of Management, Technology, and Social Sciences. 10.47992/IJMTS.2581.6012.0132.

[17] Korkontzelos, Ioannis & Nikfarjam, Azadeh & Shardlow, Matthew & Sarker, Abeed & Ananiadou, Sophia & Gonzalez, Graciela. (2016). Analysis of the effect of sentiment analysis on extracting adverse drug reactions from tweets and forum posts. Journal of Biomedical Informatics. 62. 10.1016/j.jbi.2016.06.007.

[18] Sarker A, Gonzalez G. A corpus for mining drug-related knowledge from Twitter chatter: Language models and their utilities. Data Brief. 2016 Nov 23;10:122-131. doi: 10.1016/j.dib.2016.11.056. PMID: 27981203; PMCID: PMC5144647.

[19] Silva, N. F. F. D., Coletta, L. F., & Hruschka, E. R. (2016). A survey and comparative study of tweet sentiment analysis via semi-supervised learning. ACM Computing Surveys (CSUR), 49(1), 1-26. Chicago

[20] Ribeiro, L. A., Cinalli, D., & Garcia, A. C. B. (2021, May). Discovering adverse drug reactions from twitter: A sentiment analysis perspective. In 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design

(CSCWD) (pp. 1172-1177). IEEE.

[21] Asderis, G. A. (2022). Sentiment Analysis on Twitter Data, a Detailed Comparison of TextBlob and VADER.

[22] Verma, R., Chhabra, A. & Gupta, A. A statistical analysis of tweets on covid-19 vaccine hesitancy utilizing opinion mining: an Indian perspective. Soc. Netw. Anal. Min. 13, 12 (2023). https://doi.org/10.1007/s13278-022-01015-2

[23] Najafizada, M., Rahman, A., Donnan, J., Dong, Z., & Bishop, L. (2022). Analyzing sentiments and themes on cannabis in Canada using 2018 to 2020 Twitter data. Journal of Cannabis Research, 4(1), 1-14.

[24] Talahaturuson, E., Gumelar, A. B., Sooai, A. G., Sueb, S., Suprihatien, S., Altway, H. A., ... & Sulistiyani, S. (2022, September). Exploring Indonesian Netizen's Emotional Behavior Through Investment Sentiment Analysis Using TextBlob-NLTK (Natural Language Toolkit). In 2022 International Seminar on Application for Technology of Information and Communication (iSemantic) (pp. 244-248). IEEE.

[25] Shahzad, M., & Alhoori, H. (2022). Public Reaction to Scientific Research via Twitter Sentiment Prediction. Journal of Data and Information Science, 7(1), 97-124.

[26] Abednego, L., Nugraheni, C. E., & Fedora, A. (2022). Forex Sentiment Analysis with Python. International Journal of Advanced Research in Economics and Finance, 4(1), 46-55.

[27] Yin, H., Song, X., Yang, S., & Li, J. (2022). Sentiment analysis and topic modeling for COVID-19 vaccine discussions. World Wide Web, 25(3), 1067-1083.

[28] Qian, C., Mathur, N., Zakaria, N. H., Arora, R., Gupta, V., & Ali, M. (2022). Understanding public opinions on social media for financial sentiment analysis using AI-based techniques. Information Processing & Management, 59(6), 103098.

[29] Benmelech, E., Eberly, J., Papanikolaou, D., & Krieger, J. (2021, May). Private and social returns to R&D: drug development and demographics. In AEA Papers and Proceedings (Vol. 111, pp. 336-340). 2014 Broadway, Suite 305, Nashville, TN 37203: American Economic Association.

[30] Panda, B., Panigrahi, C. R., & Pati, B. (2022, September). Exploratory and Sentiment Analysis of Classified Drug Groups. In Proceedings of the 6th International Conference on Advance Computing and Intelligent Engineering: ICACIE 2021 (pp. 229-239). Singapore: Springer Nature Singapore.

[31] Sunitha, D., Patra, R. K., Babu, N. V., Suresh, A., & Gupta, S. C. (2022). Twitter sentiment analysis using ensemble based deep learning model towards COVID-19 in India and European countries. Pattern Recognition Letters, 158, 164-170.

[32] Schlegelmilch, B. B., Sharma, K., & Garg, S. (2022). Employing machine learning for capturing COVID-19 consumer sentiments from six countries: a methodological illustration. International Marketing Review.

[33] Cameron, D., Smith, G. A., Daniulaityte, R., Sheth, A. P., Dave, D., Chen, L., ... & Falck, R. (2013). PREDOSE: a semantic web platform for drug abuse epidemiology using social media. Journal of biomedical informatics, 46(6), 985-997.

[34] Phan, N., Chun, S. A., Bhole, M., & Geller, J. (2017, April). Enabling real-time drug abuse detection in tweets. In 2017 IEEE 33rd international conference on data engineering (ICDE) (pp. 1510-1514). IEEE.

[35] Daniulaityte, R., Chen, L., Lamy, F. R., Carlson, R. G., Thirunarayan, K., & Sheth, A. (2016). "When 'bad'is 'good'": identifying personal communication and sentiment in drug-related tweets. JMIR public health and surveillance, 2(2), e6327.

Edited by: Rajni Mohana
Special issue on: Sentiment Analysis and Affective computing in Multimedia Data on Social Network
Received: Jul 31, 2023
Accepted: Sep 28, 2023