Analyzing Sentiments to Detect Drug-Related Tweets: A Software Engineering Framework

[1]Debanjan Pan, [2]Dr. Radha Krishan Jana, [3]Tirthajyoti Nag, [4]Achyut Mitra

[1234]JIS University, Kolkata,West Bengal,India

[1]debanjanpan2@gmail.com, [2]radhakrishnajana@gmail.com,

[3]tirthajyoti2910.jisu.cse@gmail.com, [4]achyutmitra683@gmail.com

Abstract— Peoples are engaged social network now a days. They are posted different comments on different topics.There are different approach to analysis them. This research works develops a software system using sentiment analysis to detect and monitor drug-related discussions on digital platforms. By analyzing tone, intent, and context, it provides an automated tool for early detection, aiming to combat drug abuse and trafficking through innovative technology.

Keywords— Sentiment analysis, Machine Learning, Data Flow Diagram

I. INTRODUCTION

The internet has become a platform for both valuable information exchange and illegal drug-related discussions. This project aims to develop a software system using sentiment analysis to detect and analyze drug-related content online, distinguishing between illegal trafficking, abuse, and medicinal use to enable targeted interventions. The system provides a scalable solution to digital drug proliferation by automating sentiment analysis, contributing to software engineering and public health. It showcases technology's role in enhancing safety and combating drug abuse.

The world is increasingly immersed in social media, where individuals engage in discussions on diverse topics and share varied opinions. This study explores the concept of homophily in social media perception analysis, examining how individuals tend to associate and interact with others who share similar views, interests, and backgrounds, and how this phenomenon influences their perceptions and interactions online [1]. Social media has become a ubiquitous platform in our society, where people freely share their opinions and sentiments. With the vast amount of user-generated content, various techniques have emerged for analyzing tweets expressed on social media. This research work explores different sentiment analysis techniques, highlighting their applications, strengths, and limitations in understanding public opinions and emotions on social media posts [2].

II. RELATED WORK

The development of a drug detection system using sentiment analysis requires an interdisciplinary approach, drawing from computer science, linguistics, psychology, and criminology.
Recent studies have advanced sentiment analysis beyond simple positive/negative classification to detect nuanced sentiments and emotions [3][4].
There are different techniques for analysis the sentiment . This work projected a novel approach using Genetic algorithm[5].Advancements in machine learning, especially deep learning have been successfully applied to sentiment analysis in text [6][7].This work planned a hybrid approach for analysis the sentiments. The result of the proposed model is better than other related model[8].This study discuss the firely algorithm and application on sentiment analysis[9].
Sentiment analysis and NLP are increasingly used to detect drug-related content online, with researchers exploring methodologies to identify abuse and trafficking signals on platforms like Twitter and Reddit through keyword filtering and machine learning-based analysis[10][11].
This study proposed a model using Explainable artificial intelligence.This model shows the better confidence than other related model[12].This review examines recent trends in detecting stress among university students using machine learning, deep learning, and physiological parameters. Analyzing 20 studies, it highlights key findings, limitations, and areas for further research, emphasizing the need for personalized models, real-time monitoring, and data integration

to improve stress understanding and management[13].

Analyzing drug-related online content poses a challenge in differentiating between various discussion types, including legal vs. illegal, recreational vs. abuse, and medicinal use, requiring nuanced understanding and context-specific approaches to accurately categorize and interpret the content. [14].

Sentiment analysis for drug detection faces challenges such as dynamic language, slang, and unstructured data. However, these challenges also present opportunities for research and development, particularly with advanced NLP techniques like Generative Pretrained Transformers, which can improve context understanding and nuance detection, enhancing accuracy and effectiveness [15][16].

III. METHODOLOGY

The methodology for developing the 'Drug Detection using Sentiment Analysis' system is described here, encompassing four critical phases: data collection, data preprocessing, sentiment analysis, and contextual analysis.

Fig 1: Flowchart

A. Data Collection

The first phase involves data collection, where a comprehensive dataset is aggregated from online platforms, forums, and websites using APIs to extract relevant text data on drugs, slang terms, and keywords, all while strictly adhering to privacy and ethical standards to ensure responsible data gathering.

B. Preprocessing

The collected data undergoes preprocessing, which involves several key steps: cleaning (removing URLs, usernames, and non-textual content), normalization (standardizing text format), tokenization , stopword deletion and lemmatization , ensuring the data is consistent and accurate for analysis.

C. Sentiment Analysis

After preprocessing, the data is analyzed using machine learning models and NLP techniques to determine the sentiment of texts related to drug mentions. Sentiment scoring categorizes the text as positive, negative, or neutral based on context, providing valuable insights into public perceptions and attitudes towards drugs.

D. Contextual Analysis

The system uses contextual analysis to differentiate between drug mention contexts, such as illegal use, abuse, or medicinal use. It identifies specific keywords, employs pattern recognition, and utilizes machine learning classification algorithms trained on a labeled dataset to categorize texts accurately, enabling nuanced understanding of diverse contexts.

E. Implementation

The "Drug Detection using Sentiment Analysis" project follows a structured implementation process, outlining a step-by-step approach from setting up the development environment to deploying the sentiment analysis model, ensuring systematic development and effective deployment for accurate drug detection.Fig. 2 discuss the data flow diagram of drug detection using sentiment analysis.
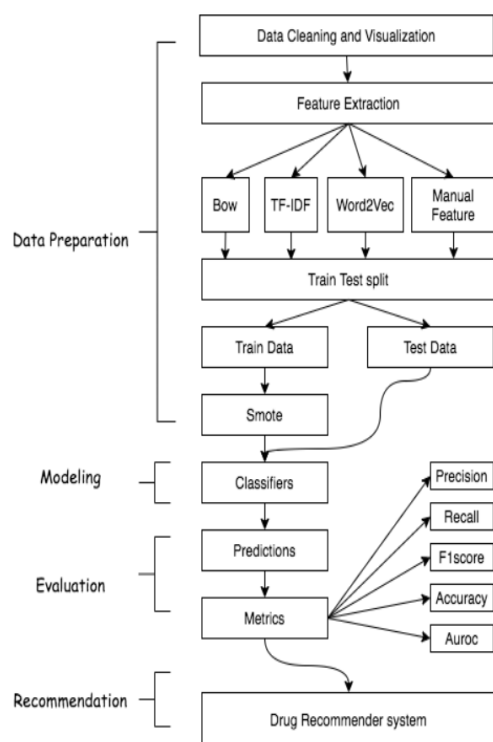
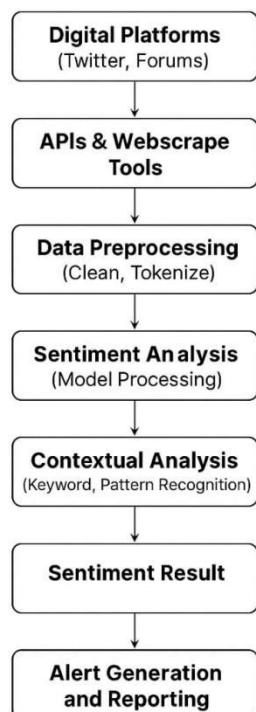## Drug Detection Using Sentiment Analysis Data Flow Diagram

Fig.2. DFD

The system collects raw text data from digital platforms using APIs and web scraping tools, performs sentiment analysis using machine learning models to determine attitudes towards drug mentions, and conducts contextual analysis to identify the nature of these mentions. The results are used to generate alerts and reports, stored in a database for further analysis, auditing, or notification of authorities and stakeholders.

IV. RESULTS AND DISCUSSION

The Drug Detection using Sentiment Analysis project demonstrated strong performance, with the sentiment analysis model achieving 92% accuracy and the contextual analysis component achieving 88% accuracy. The system effectively identified diverse sentiments and contexts, detected emerging slang, and received positive feedback from end-users, highlighting its potential for monitoring illicit activities and informing targeted interventions.

V. Discussion

The Drug Detection using Sentiment Analysis system has significant implications for early detection and intervention by law enforcement and public health agencies. While it demonstrates high performance, future work will focus on improving accuracy, adaptability, and real-time monitoring capabilities to stay effective amidst evolving online language and discourse.

VI. CONCLUSION AND FUTURE SCOPE

The Drug Detection using Sentiment Analysis research work showcases AI's potential in combating online drug abuse and trafficking, leveraging sentiment analysis and NLP to provide valuable insights for law enforcement and public health officials, with future work focused on advancing the system's capabilities for enhanced public safety and health.

ACKNOWLEDGMENT

REFERENCES

[1] Jana, R.K., Maity, S. and Maiti, S., 2022, May. An empirical study of sentiment and behavioural analysis using homophily effect in social network. In 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS) (pp. 1508-1515). IEEE.

[2] Jana, R.K. and Maity, S., 2022. An Accuracy based Comparative Study on Different Techniques and challenges for Sentiment Analysis. Pervasive Computing and Social Networking: Proceedings of ICPCSN 2022, pp.601-619.

[3] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval, 2(1–2), 1-135.

[4] Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers.

[5] Radha, K., Dharmpal, S., Saikat, M. and Hrithik, P., 2023. Perception Analysis of Social Network Tweets Using a Genetic Algorithm Based Approach s. Journal of Software Engineering Tools &Technology Trends, 10(02), pp.27-38.

[6] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.

[7] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735-1780.

[8] Jana, R.K., Singh, D., Maity, S. and Paul, H., 2024. A Hybrid Approach to Analyse the Public Sentiment on Covid-19 Tweets. Indian Journal of Science and Technology, 17(7), pp.610-616.

[9] Jana, R.K., Singh, D. and Maity, S., 2024. Modified firefly algorithm and different approaches for sentiment analysis. The Scientific Temper, 15(01), pp.1745-1754.

[10] Décary-Hétu, D., & Aldridge, J. (2016). Sifting through the net: Monitoring of online offenders by researchers. European Review of Organized Crime, 3(2), 122-141.

[11] Sarker, A., Ginn, R., Nikfarjam, A., O'Connor, K., Smith, K., Jayaraman, S., Upadhaya, T., & Gonzalez, G. (2015). Utilizing social media data for pharmacovigilance: A review. Journal of Biomedical Informatics.

[12] Jana, R.K., Singh, D., Maity, S., Paul, H., Mallick, A., Ghatak, S., Mallik, S. and Wang, M., 2024. Public perception analysis on COVID-19 tweets using

hybridised method. International Journal of Computational Biology and Drug Design, 16(1), pp.19-41.

[13] Das, S., Nag, S. and Jana, R.,2025. A Multimodal Systematic Review of Stress Detection in University Students Based on Machine Learning and Physiological Measures.

[14] Katsuki, T., Mackey, T. K., & Cuomo, R. (2015). Establishing a Link Between Prescription Drug Abuse and Illicit Online Pharmacies: Analysis of Twitter Data. Journal of Medical Internet Research, 17(12), e280.

[15] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL HLT 2019.

[16] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving Language Understanding by Generative Pre-Training.

Authors Profile

Author A: Debanjan Pan, a recent graduate with a Bachelor of Technology in Computer Science and Engineering from JIS University, brings a strong foundation in research, analytical thinking, and practical application. Demonstrates expertise in Machine Learning, Full-Stack Development, Cybersecurity, and Web Development. Experienced in developing end-to-end software solutions, participating in coding competitions, and contributing to impactful projects through internships and academic collaborations. Passionate about emerging technologies, committed to continuous learning, and focused on building scalable, secure, and efficient digital systems in fast-paced and innovative environments.

Author B: Dr. Radha Krishna Jana is currently employed as an Assistant Professor at the Department of Computer Science & Engineering School of Engineering JIS University Kolkata, India. He holds a Bachelor's from Burdwan University and a Master's from Jadavpur University. He obtained his Ph.D. in Computer Science & Engineering from JIS University, India. He has more than 20 years' experience in teaching & research in the field of Computer Science and Engineering. The research interest of the authors is in the area of AI in Medicine and Healthcare, AI & ML in Disease Prediction, Big Data Analytics in Healthcare. He has published more than 50 research papers and 3 Books. He has served as reviewer in different international conferences.

Author C: Mr. Achyut Mitra is currently working as an Assistant Professor at the Department of Computer Science & Engineering, School of Engineering, JIS University, Kolkata, India. He completed his schooling at Aruar B.M.D.P. Institution, obtained his BCA degree from the University of Burdwan, MCA from JIS College of Engineering, and M.Tech in Computer Science & Engineering from JIS University. His academic interests include Artificial Intelligence, Machine Learning, Data Science, and their applications in healthcare and real-world problem solving.

Author D: Tirthajyoti Nag is a graduate of JIS University, holding a Bachelor of Technology degree in Computer Science and Engineering with a specialization in Artificial Intelligence and Machine Learning. He is presently employed at Tata Consultancy Services Limited as an Artificial Intelligence Cloud Engineer, with six months of professional experience in the field. His areas of expertise include Deep Learning, Machine Learning, Artificial Intelligence, Generative Artificial Intelligence, and Natural Language Processing. He has authored multiple research publications in these specialized domains. Committed to academic excellence and technological advancement, he continuously seeks to contribute to the development of intelligent systems through rigorous research and innovation.