DRUGS OVERDOSE DEATHS IN USA

Mohana Satyanarayana Moganti,
San Jose State University, CA

Supriya Katragadda,
San Jose State University, CA

Naga Varun Bathina,
San Jose State University, CA

David Wang
San Jose State University, CA

1. ABSTRACT

In the 1990s, major pharmaceutical companies launched a series of potent painkillers, which were widely prescribed by physicians to alleviate patients' discomfort, unknowingly sparking a nationwide addiction crisis and a sharp increase in drug-related deaths, mainly due to serious side effects like breathing difficulties and mood swings. Surprisingly, this essential issue has received insufficient attention, and complete solutions remain scarce.

In response, we have crafted a detailed analytical model utilizing the meticulously cleaned VSRR Provisional dataset, aiming to forecast the incidence of overdose-related fatalities, discern patterns in drug-specific deaths over time, and project the likely scenario for 2024. Our objectives are to reduce the effect of this ongoing catastrophe, save lives, and give communities with the resources they need for prevention, as well as to increase public knowledge of the possible risks associated with these substances.

2. INTRODUCTION

In recent times, the U.S. has observed a striking surge in drug overdose-related fatalities. This health quandary impacts not only people, but also communities and society as a whole. Innovative data analysis and data-centric techniques have proven critical in understanding and reducing drug overdoses as we negotiate the complexity of this opioid crisis. Using data from the Centers for Disease Control and Prevention (CDC), the proposed model focuses on developing a predictive framework for drug overdose. This CDC repository provides detailed information on drug overdose events, including spatial and temporal patterns, impacted individual profiles, and chemicals involved. Our endeavour aims to achieve numerous critical goals by leveraging the power of data analytics and current algorithmic methodologies. Drawing from the exhaustive records in the CDC collection, our initiative aims to lay a data-backed groundwork to confront this pressing public health challenge. While predictive algorithms are not a complete solution, we recognize that they can significantly improve resource distribution and foresight, contributing to a reduction in the harsh consequences of drug overdoses in our communities. The United States has recently seen a dramatic increase in drug overdose-related fatalities. This health quandary affects not only individuals, but also communities and society as a whole. Innovative data analysis and data-centric strategies have proven critical in understanding and reducing drug overdoses

as we navigate the complexities of this opioid crisis. Using data from the Centers for Disease Control and Prevention (CDC), the proposed model focuses on developing a predictive framework for drug overdose. This CDC repository offers rich insights into drug overdose incidents, capturing spatial and temporal trends, affected individual profiles, and involved substances. By channelling the prowess of data analytics and modern algorithmic techniques, our endeavour targets several pivotal goals. Drawing from the exhaustive records in the CDC collection, our initiative aims to lay a data-backed groundwork to confront this pressing public health challenge. We recognize that while predictive algorithms are not a complete solution, they can substantially optimize resource distribution and foresight, contributing to a reduction in the harsh consequences of drug overdoses in our communities.

3. JUSTIFICATION

The model offers a promising avenue for influencing health guidelines, initiatives, and improving the lives of communities grappling with the opioid dilemma. It stands as an instrumental tool for decision-makers, policymakers, and medical experts aiming to address the severe outcomes of drug overdoses in the United States. This initiative aids in the swift recognition of potential risks, paving the way for the prompt execution of preventive measures. Moreover, it provides data-driven perspectives intended to decrease overdose occurrences and formulate sustainable strategies to counteract drug overdose challenges.

In essence, the project empowers professionals in healthcare to employ insights from the predictive model, enabling specialized counselling, comprehensive treatment accessibility, and subsequently diminishing the prospects of overdose events.

4. DATASET OVERVIEW

Data was collected from The Centres for Disease Control and Prevention (CDC) provided in the form of VSRR preliminary Drug Overdose Death Counts. The information spans the years 2015 through 2023 and includes information on several states as well as the kind of substance that resulted in the death. And the total data is updated on a monthly basis, encompassing a range of medications and states.

Overall, there are 60601 data points from the states of Alaska, Alabama, Arkansas, Arizona, California, Colorado, Indiana, and a few additional places, spanning from April 2015 to May 2023. In the dataset, the different drug categories included include Cocaine (T40.5), Heroin (T40.1), Opioids (T40.0-T40.4, T40.6), Methadone (T40.3).

4.1. Memory Footprint

The entire dataset occupies approximately 5.5 MiB in memory, with an average record size of 96.0 B. This is a moderate size, indicating that the dataset should be manageable on most modern computing environments without necessitating special hardware or distributed computing resources.

4.2. Variable Types

The dataset includes a diverse set of variable types:

Text Variables: There are 2 text variables. These could represent strings, categories, or any other form of textual data. Text data can be crucial for capturing qualitative attributes and might require special pre-processing steps such as tokenization, embedding, or vectorization for further analysis.

Numeric Variables: 4 variables are numeric. These could represent interval or ratio scale data and are suitable for various mathematical and statistical operations. Numeric variables are crucial for regression

analyses, correlation studies, and other quantitative research methods.

Categorical Variables: There are 6 categorical variables. These variables represent categories and are essential for classification tasks, group analyses, and summarizing data through grouping. Categorical variables may need to be encoded or transformed if they are to be used in machine learning models.

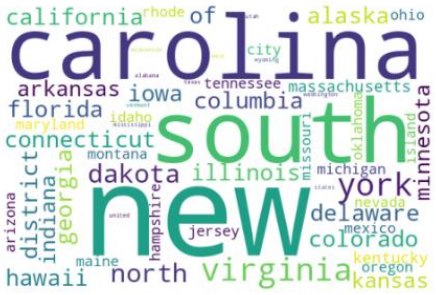

Fig.1. State WordMap

4.3. Data Quality

The dataset exhibits a 5.6% rate of missing values, totalling 40,763 missing cells. This is a significant aspect of data quality and requires attention during the pre-processing stage of data analysis. Strategies to handle these missing values might include imputation, removal of affected rows or columns, or using algorithms that can handle missing data.

There are no duplicate rows present in the dataset, accounting for 0% of the data. This is a positive indicator of the data's integrity, reducing the need for deduplication processes and ensuring that analyses are not skewed by redundant information.

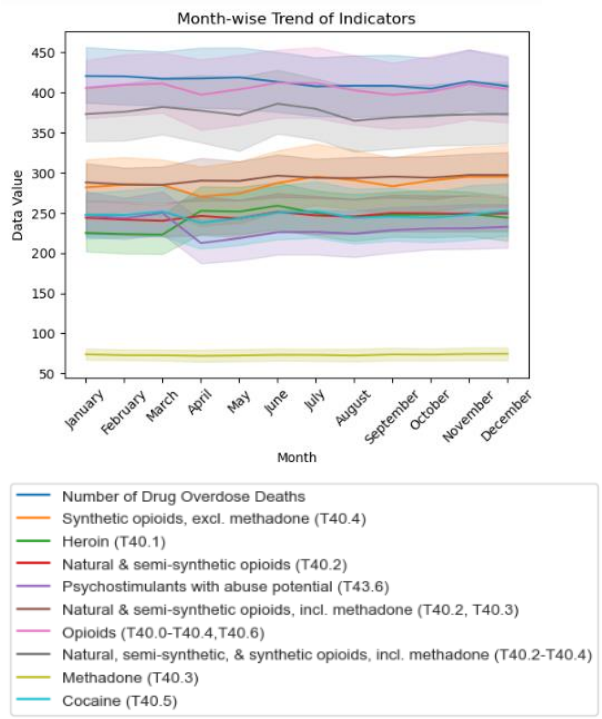

Fig.1.1. Month-wise Trend

5. DATA ANALYSIS

In the course of our data analysis, we observed several significant characteristics and patterns that are crucial for understanding the dataset and subsequently making informed decisions for any predictive modelling or statistical analysis.

5.1. Constant Values and High Correlation

The *'Period'* variable has been found to hold a constant value throughout the dataset, suggesting a lack of variance that could impact its utility in predictive modelling.

A substantial correlation is evident between *'Year'* and *'Footnote Symbol'*, highlighting a potential dependency between these two variables.

Similarly, *'Data Value'* and *'Predicted Value'* exhibit a high overall correlation, suggesting that these variables share a substantial amount of variance. This can be crucial for

regression tasks as multicollinearity could impact the performance of the models.

The *'Percent Complete'* also shows a high correlation with *'Footnote Symbol'*, which is an essential aspect to consider for data pre-processing.

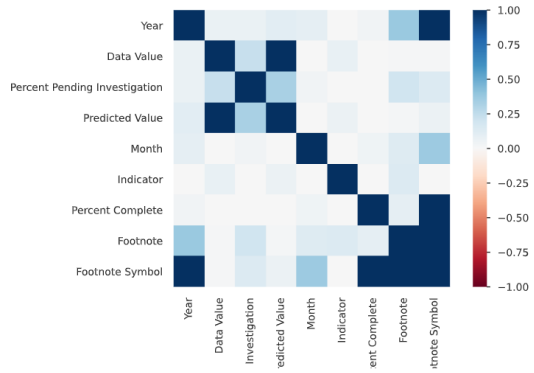

Fig.2. HeatMap

5.2. Imbalances and Missing Values

The dataset presents significant imbalances, particularly in *'Percent Complete'* (99.7%), *'Footnote'* (57.8%), and *'Footnote Symbol'* (51.1%). Such imbalances could lead to biased models, especially if the target variable is correlated with these features.

Missing values are another concern, with *'Data Value'* missing 18.2%, *'Footnote'* missing 6.7%, *'Footnote Symbol'* missing 6.9%, and *'Predicted Value'* missing a substantial 35.4% of its data. Addressing these missing values is crucial for maintaining the integrity of any analytical model.

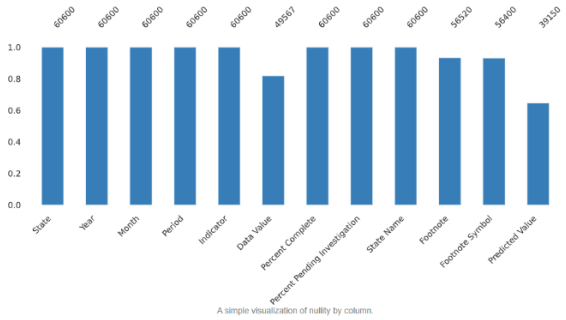

Fig.3. Missing value graph

5.3. Skewness and Zeros:

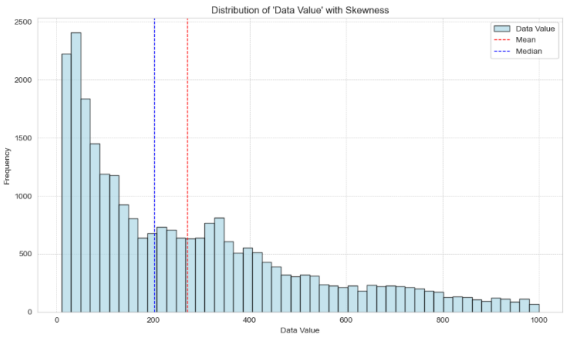

Fig.4. Distribution of Data Value

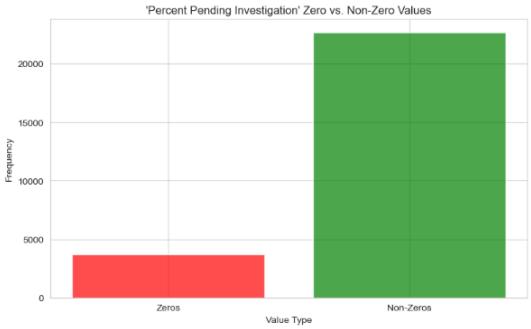

Fig.5. Zero's and Non-Zero's for Percent Pending Inv

The *'Data Value'* variable is highly skewed ($\gamma 1 = 21.25519312$), indicating a distribution that is far from normal. Depending on the nature of the subsequent analysis, this could necessitate a transformation to approximate a normal distribution.

Additionally, *'Percent Pending Investigation'* has 12.5% of its values as zeros, which could be indicative of either a placeholder value or a genuine zero, both of which would require different handling strategies.

5.4. Recommendations for Pre-processing:

a. Addressing Constants and High Correlation: Variables like *'Period'* which are constants can be removed as they do not contribute to model variance. For highly correlated variables, consider using dimensionality reduction techniques like Principal Component Analysis (PCA) or removing one of the correlated variables to reduce multicollinearity.

b. Handling Imbalanced Data: Techniques like oversampling the minority class, under sampling the majority class, or using advanced algorithms that are robust to class imbalance can be employed.

c. Dealing with Missing Values: Depending on the nature of the missingness, strategies such as imputation, deletion, or employing algorithms that can handle missing values should be considered.

d. Transforming Skewed Data: Applying transformations like log, square root, or Box-Cox transformations can help in normalizing the distribution of skewed variables.

e. Analysing Zeros: Investigate the nature of zeros in 'Percent Pending Investigation' to decide on the appropriate handling method.

By addressing these issues appropriately, we can ensure that the dataset is well-prepared for any subsequent analysis or modelling, leading to more reliable and valid results.

6. RELATED WORK

The drug overdose crisis, especially related to opioids, has been a subject of intense study and intervention over the past decade. Researchers, policymakers, and healthcare professionals have been working tirelessly to understand the dynamics of drug use and its subsequent effects on public health, aiming to develop effective strategies for prevention, treatment, and harm reduction. Below is a detailed review of related work in this domain, drawing parallels and distinctions with our current project.

Epidemiological Studies and Trends Analysis

Numerous studies have provided extensive epidemiological analysis of drug overdose trends, identifying key demographic factors, geographic hotspots, and temporal patterns associated with overdose deaths. The work by Friedman et al. (2023) and the reports from the National Institutes of Health (2023) have played a crucial role in mapping out the landscape of the drug overdose crisis, setting a foundational understanding upon which further research can build. Our project extends this line of work by not only analysing past and present trends but also employing predictive modelling to forecast future occurrences of drug overdose deaths. This provides an additional layer of utility for policymakers and healthcare providers, offering foresight to better allocate resources and implement preventative measures.

Machine Learning and Predictive Modelling

In the realm of machine learning, several studies have attempted to predict drug overdose occurrences and outcomes using various algorithms and datasets. The application of predictive modelling in this context holds immense potential to augment human decision-making, providing data-driven insights to guide interventions. For instance, a study conducted by Yousif (2023) utilized machine learning techniques to identify individuals at high risk of opioid overdose, helping to prioritize interventions for those most in need. While our project also employs machine learning for predictive purposes, our focus is broader, aiming to predict drug overdose deaths on a population level using publicly available data from the CDC. This aligns with the approach taken in the work by the Centers for Disease Control

and Prevention (2023b), although our methodology and specific objectives may differ.

Community Engagement and Public Health Interventions

Addressing the drug overdose crisis requires more than just understanding the problem; it necessitates active intervention and community engagement. Numerous initiatives have been launched to raise awareness, provide support, and implement preventative measures at the community level. The integration of community policing, as highlighted in our community contribution section, is reminiscent of the approaches discussed in various literature, emphasizing the importance of multi-faceted strategies to combat drug overdose. Furthermore, the distribution of naloxone and the establishment of safe disposal sites, as proposed in our project, have been widely advocated for and implemented in various regions, showcasing their effectiveness in harm reduction.

Limitations and Future Work

Despite the strides made in understanding and addressing the drug overdose crisis, there remain significant gaps and challenges. The limitations in data quality, such as missing values and imbalances, as detailed in our data analysis section, are not unique to our dataset. These issues pervade many public health datasets, necessitating careful pre-processing and validation of predictive models. Future work in this domain must continue to refine data collection and processing techniques, ensuring that the models developed are robust, reliable, and applicable across diverse contexts.

7. COMMUNITY CONTRIBUTION

Overdose, particularly from illicit drugs, presents grave challenges to society. Beyond the immediate threat to life, it fosters a chain of negative repercussions that diminish the overall quality of community living. Issues stemming from overdose, such as unemployment and deteriorating mental health, necessitate concerted community actions to address the epidemic. This document delves into the crucial contributions that can be initiated by communities to combat the overdose crisis.

7.1. Awareness and Education: Elevating community knowledge is paramount. Organizing detailed workshops, comprehensive seminars, and widespread informational campaigns can enlighten individuals about the perils of drug misuse, thereby empowering them to make informed decisions.

7.2. Enhanced Support Groups: The emotional aftermath of drug overdose isn't limited to the affected individuals alone; families too bear the brunt. By establishing robust support groups, communities can offer both emotional solace and pragmatic guidance to affected parties.

7.3. Strengthened Advocacy: A proactive community can champion the cause by advocating for improved access to vital services. This includes lobbying for comprehensive addiction treatments, addressing underlying mental health challenges, and promoting strategies that reduce harm.

7.4. Tailored Prevention Programs: Prevention remains better than cure. By launching initiatives explicitly aimed at the youth, educational institutions, and families, communities can spotlight the hazards of drugs and concurrently advocate for healthier living choices.

7.5. Community Policing Collaboration: A holistic approach to the overdose crisis includes tackling its root causes. Collaborative efforts with law enforcement agencies can stymie drug trafficking, significantly curtailing the influx of harmful substances into communities.

7.6. Accessible Treatment Facilities: Recovery begins with accessibility. Communities can bolster recovery efforts by either supporting existing treatment centers or establishing new ones, ensuring that individuals grappling with addiction have the resources they need within arm's reach.

7.7. Naloxone Distribution Training: Empowering community members with the knowledge and resources to administer naloxone—an opioid overdose reversal medication—can be the difference between life and death.

7.8. Safe Medication Disposal Sites: By setting up designated disposal hubs for unused or expired medications, communities can preclude potential misuse and reduce the environmental hazards posed by improper disposal.

7.9. Engaging Media Campaigns: Harnessing the power of local media to share both triumphant stories of recovery and the grim aftermath of drug misuse can sway public sentiment and foster a community-driven desire for change.

7.10. Holistic Counselling Services: Addiction doesn't exist in a vacuum. By offering comprehensive counselling services, communities can address both the addiction itself and the underlying triggers, promoting holistic healing.

7.11. Prompt Crisis Intervention: In moments of extreme vulnerability, immediate assistance is crucial. Establishing dedicated helplines or crisis intervention services ensures that those in dire need receive timely and appropriate aid.

Addressing the overdose epidemic demands a multi-pronged, community-centric approach. By implementing the aforementioned contributions, communities can not only mitigate the current crisis but foster an environment conducive to long-term recovery and holistic well-being.

8. CONCLUSION

The opioid epidemic presents a critical public health challenge, marked by a significant rise in drug overdose-related deaths. The proposed model utilizing the VSRR Provisional dataset from the CDC aimed to leverage machine learning techniques to predict drug overdose deaths, providing crucial insights for preventative measures and resource allocation. Our comprehensive data analysis revealed patterns, correlations, and data quality issues essential for understanding the epidemic's dynamics and informing our predictive modeling efforts.

Addressing the opioid epidemic requires a multifaceted approach, and the predictive model proposed in this report stands as a significant step towards understanding and mitigating drug overdose deaths. By leveraging comprehensive data and advanced machine learning techniques, the model provides valuable insights that can guide interventions, shape policies, and ultimately contribute to saving lives.

However, it is crucial to acknowledge the limitations and challenges, such as data imbalances and missing values, and take proactive steps to address these issues during the data pre-processing and model development stages. With a careful and informed approach, the predictive model has the potential to become a vital tool in the fight against the opioid crisis, fostering safer communities and improving public health outcomes.

In conclusion, this project represents a critical step towards understanding and mitigating the drug overdose crisis through predictive analytics, data-driven insights, and community engagement. While challenges remain, the potential of such approaches to save lives, inform policy, and contribute to the overall well-being of communities is immense, underscoring the

urgent need for continued research, innovation, and action in this domain.

9. REFERENCES

i. Centers for Disease Control and Prevention. (2023a, January 5). *NCHS Data Visualization Gallery - drug poisoning mortality*. Centers for Disease Control and Prevention. https://www.cdc.gov/nchs/data-visualization/drug-poisoning-mortality/

ii. Centers for Disease Control and Prevention. (2023b, October 11). *Products - vital statistics rapid release - provisional drug overdose data*. Centers for Disease Control and Prevention. https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.html

iii. *Drug overdose death statistics [2023]: Opioids, fentanyl & more*. NCDAS. (2023, January. https://drugabusestatistics.org/drug-overdose-deaths/#:~:text=Over%2096%2C700%20people%20die%20from%20drug%20overdoses%20in%20a%20year.&text=Opioids%20are%20a%20factor%20in%207%20out%20of%20every%2010%20overdose%20deaths.&text=Drug%20overdoses%20have%20killed%20almost%20a%20million%20people%20since%201999 .

iv. U.S. Department of Health and Human Services. (2023, September 25). *Drug overdose death rates*. National Institutes of Health. https://nida.nih.gov/research-topics/trends-statistics/overdose-death-rates

v. Yousif, N. (2023, September 17). *How the fentanyl crisis' fourth wave has hit every corner of the US*. BBC News. https://www.bbc.com/news/world-us-canada-66826895

vi. Drug Overdose Deaths Involving Cocaine and Psychostimulants with Abuse Potential — United States, 2003–2017. *Published online 2019 May 3.* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6541315/

vii. Trends in **Drug Overdose** Deaths Among US Adolescents, January 2010 to June 2021 - *Friedman J, Godvin M, Shover CL, Gone JP, Hansen H, Schriger DL.* https://pubmed.ncbi.nlm.nih.gov/35412573/

viii. ydata-profiling a python library. https://pypi.org/project/ydata-profiling/