Dynamics of Substance Abuse A Data-Driven Approach to Public Health Strategy and Economic Influence

Jaimin Rakesh Shah
Goergen Intitute for Data Science
University of Rochester
Rochester, New York
jshah15@ur.rochester.edu

Shyam Chirag Shah
Goergen Intitute for Data Science
University of Rochester
Rochester, New York
sshah77@ur.rochester.edu

***Abstract*— This study addresses the multifaceted challenge of substance abuse, a significant public health issue with complex determinants including socioeconomic factors, demographic variables, and regional differences. Through the analysis of over 2.2 million records, we aim to discern patterns and trends in substance abuse, specifically examining the impact of the 2008 economic crisis on usage patterns and the predictive potential of economic indicators on public health outcomes. The research employs advanced data mining techniques such as time series analysis with Prophet and pattern recognition to provide an in-depth analysis of substance abuse evolution. We also explore the efficiency of various classifiers – Random Forest, CatBoost, and Naive Bayes – in predicting hospital substance abuse cases. Our findings highlight the prevalence of alcohol abuse among individuals under 21, particularly within the 'Young' demographic, and significant drug use within the 'Middle Age' group, informing targeted intervention strategies. Jaccard Similarity Heatmaps reveal both shared and unique substance abuse patterns across metropolitan areas, suggesting the need for tailored public health responses. Additionally, feature importance analysis demonstrates that factors of local importance may not hold the same weight on a larger scale. This study offers a comprehensive overview of substance abuse trends and provides insights critical for developing informed, effective public health strategies.**

Keywords*— *Data visualization, Time series analysis, Pattern recognition, Prophet forecasting, Random Forest Algorithm, CatBoost Classifier, Naive Bayes Classifier, Jaccard Similarity Heatmap, Granger causality test, FP-Growth algorithm, Predictive modelingIntroduction.

I. Introduction

In the intricate landscape of public health, the issue of substance abuse stands as a formidable challenge, posing significant implications for policymaking and resource allocation. The complex and dynamic nature of drug abuse patterns, influenced by a myriad of factors such as age, race, and geography, presents a daunting task for accurate prediction and intervention. This research paper delves into the depths of this pervasive issue by leveraging the power of advanced data mining techniques. Employing methods like time series analysis and pattern recognition, our study seeks to unravel the intricate web of factors contributing to substance abuse. We analyze a dataset comprising over 2.2 million records across multiple variables to provide a comprehensive overview of the trends and patterns in substance abuse.

This paper explores the impact of various socioeconomic factors, including the 2008 economic crisis and its potential correlation with substance use patterns across different regions. Notably, we investigate whether the US Housing Index crash led to a dip in substance abuse rates. We utilize sophisticated models, including Prophet, to understand and predict the trajectory of this pressing public health issue. Additionally, we compare the performance of three classifiers – Random Forest Algorithm, CatBoost Classifier, and Naive Bayes Classifier – to build a model that predicts the type of substance abuse in hospitals. Our research offers valuable insights into the dark patterns of drug abuse, paving the way for more informed and effective strategies to combat this ongoing crisis.

II. Literature Survey

"The Impact of the 2008 Economic Crisis on Substance Use Patterns in the Countries of the European Union" examines the nuanced effects of the 2008 economic crisis on substance use. It finds that, generally, the crisis led to a downtrend in substance use in the broader EU population. This trend is attributed to reduced disposable income, leading to decreased consumption of alcohol and tobacco. However, among specific vulnerable groups, particularly those affected by job loss and prolonged unemployment, there was an increase in harmful substance use behaviors, including binge drinking and illicit drug use. The study substantiates these findings through various data sources and studies across EU countries, revealing mixed effects on alcohol consumption, an increase in smoking prevalence in Italy, and a shift towards cheaper illicit drugs. The review underscores the crisis's complex impact, highlighting both a general decrease in substance use and an increase among certain subgroups, stressing the need for focused public health strategies during economic downturns.[4]

In his 1998 study, "Drug Abuse-Related Mortality in the United States: Patterns and Correlates," Jeffrey E. Kallan examines the sociodemographic factors influencing psychoactive drug-related mortality (PDM) in the U.S. Utilizing data from the National Health Interview Survey linked with the National Death Index, the study identifies significant determinants of PDM, including age, sex, race, marital status, income, and health status. Kallan's analysis highlights that risk varies notably across different demographic groups, with black males showing a particularly high risk. The study also underscores the impact of socioeconomic factors and poor perceived health status on drug-related mortality, while acknowledging limitations like the inability to differentiate drug types or intentions behind drug use. This research underscores the complex interplay of various sociodemographic factors in drug-related deaths, pointing to the need for targeted interventions considering these disparities.[1]

The Executive Summary of the World Drug Report 2018 highlights key global trends and challenges in drug use and trafficking. It reports that around 275 million people worldwide, or 5.6% of the global population aged 15–64 years, used drugs at least once in 2016. Approximately 31 millions of these individuals suffer from drug use disorders. Opioids remain the most harmful drug type, implicated in 76% of deaths associated with drug use disorders. The report also notes significant geographical variations in drug use: opioid use is a major concern in North America and parts of Africa and Asia, while non-medical use of prescription drugs is also rising globally. Additionally, the report discusses the expansion of drug markets, highlighting the coexistence of traditional drugs like heroin and cocaine with new psychoactive substances and the growing non-medical use of prescription drugs. It calls for enhanced responses to these evolving challenges, emphasizing the need for comprehensive, evidence-based strategies to tackle both supply and demand aspects of the global drug problem.[2]

III. Objective

The central objective of this research is to provide an in-depth analysis of the evolution and dynamics of drug abuse, a critical issue in public health. Our first aim is to dissect the complex web of factors contributing to the changing landscape of substance abuse. This exploration extends to investigating the intricate relationship between drug abuse and economic conditions, with a particular focus on the impact of significant economic events like the 2008 Housing Index Crash.

In parallel, our research is dedicated to identifying distinct patterns of substance abuse in various metropolitan areas across the United States. This aspect of the study is geared towards uncovering region-specific trends and factors. In parallel, our research is dedicated to identifying distinct patterns of substance abuse in various metropolitan areas across the United States. This aspect of the study is geared towards uncovering region-specific trends and factors.

IV. Methodology and Implementation

A. Visualizations

The initial stage of our data analysis involved a comprehensive examination of various visual representations of drug abuse cases across the United States. These visualizations were instrumental in identifying key patterns and trends which informed our data pre-processing strategy.

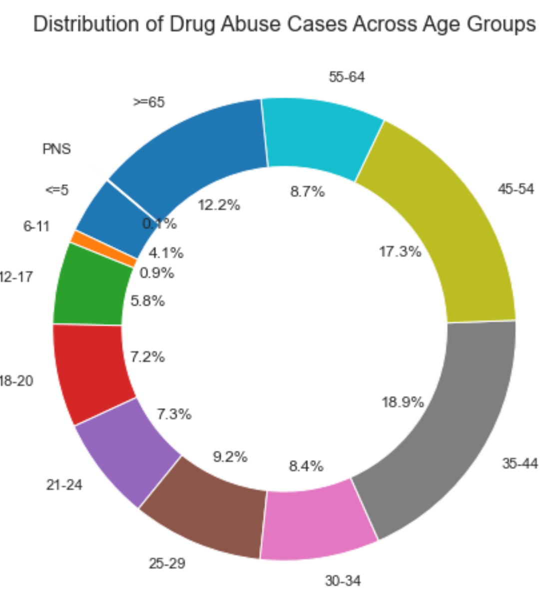

Fig. 1 Distribution of Drug Abuse Cases Across Age Groups

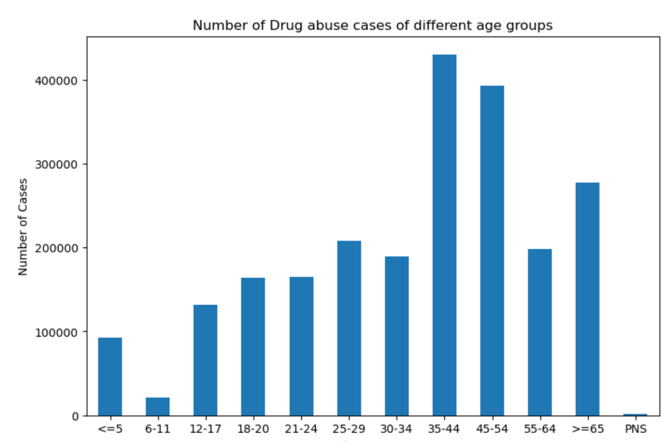

Fig. 2 Number of Drug Abuse Cases of Different Age Groups

Age distribution of drug abuse cases, depicted by a donut chart and a bar chart, revealed a pronounced prevalence within the 25-54 age range, peaking at 35-44 years. This discovery necessitated the categorization of age data into policy-relevant cohorts, facilitating targeted analysis of demographic groups most affected by substance abuse.

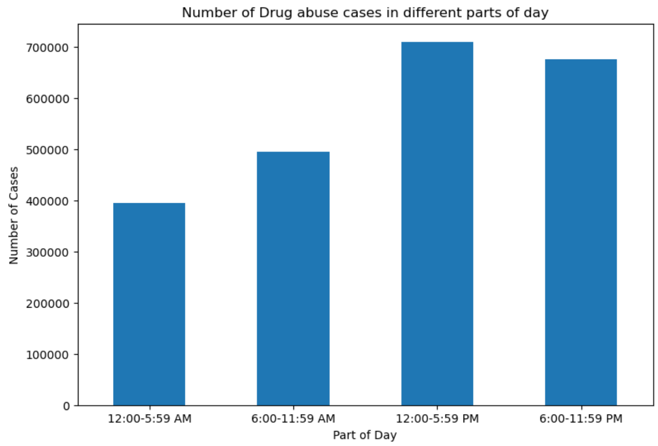

Fig. 3 Number of Drug Abuse Cases in Different Parts of Day

Diurnal variations, indicated by a higher frequency of cases in the afternoon and evening hours, were noted. This pattern influenced the planning of interventions and may also direct the scheduling of public health services and resource allocation.

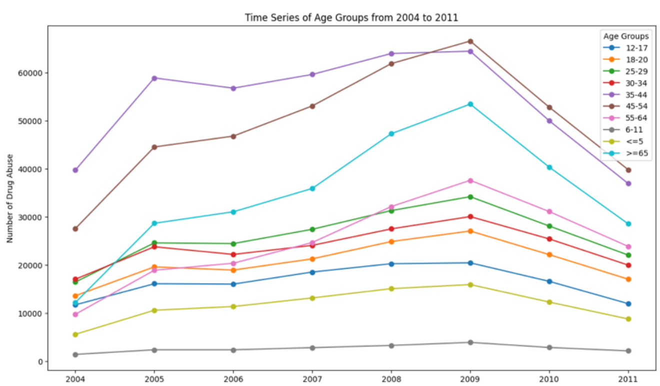

Fig. 4 Time Series of Age Groups from 2004 - 2011

The longitudinal analysis of drug abuse cases segmented by age groups over an eight-year period provides a comprehensive overview of the shifting landscape of substance abuse within different age demographics. The time series graph offers a visualization of trends that are critical to understanding the evolving dynamics of drug abuse. This visualization can be instrumental in assessing the long-term impact of public health interventions and policy changes targeted at various age groups. Additionally, it may reveal emerging patterns of drug use among specific age cohorts, which could reflect broader societal shifts, the introduction of new substances, or changes in drug availability. By tracking these trends over time, researchers and policymakers can identify age groups that are increasingly susceptible to drug abuse, potentially guiding the allocation of resources and the development of age-specific educational and prevention programs.

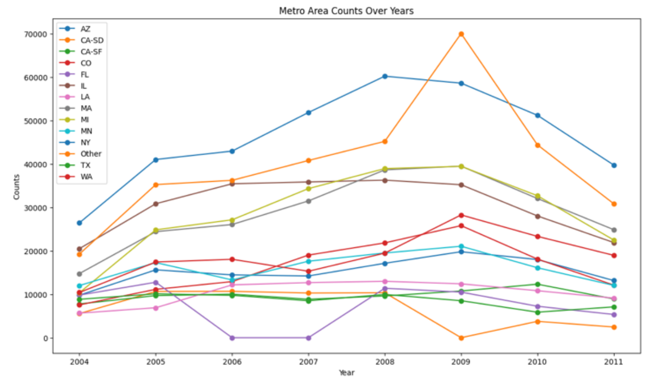

Fig. 5 Time Series of Number of Cases in Metros

The spatial distribution of drug abuse cases in major metropolitan areas over the same period highlights the influence of urban environmental factors on substance abuse trends. The "Metro Area Counts Over Years" graph delineates the ebb and flow of drug abuse incidents across densely populated regions, suggesting a correlation between urbanization and drug-related activities. This pattern underscores the importance of considering the unique challenges and characteristics of metropolitan areas when designing and implementing drug prevention and treatment strategies. Factors such as economic conditions, law enforcement policies, and the availability of healthcare resources within these urban centers can have a significant impact on the observed trends. The visualization aids in identifying metropolitan areas with particularly high or increasing counts of drug abuse cases, which could be indicative of a need for intensified intervention efforts or further study to understand the underlying causes.

B. *Time Series Forecasting*

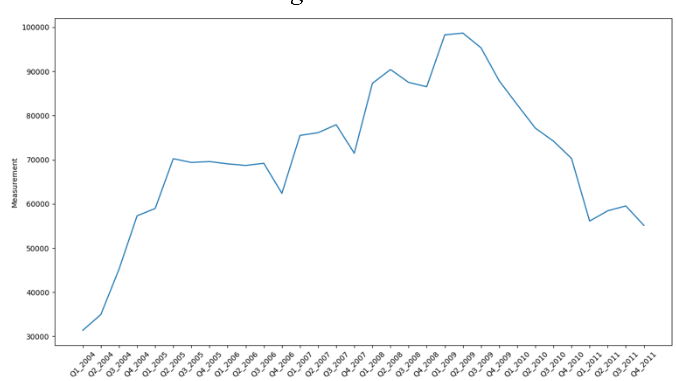

Fig. 6 Time Series of cumulative data

The time series chart of drug abuse cases from 2004 to 2011 presents an intriguing narrative of fluctuation and trend reversal, illustrating the complexity inherent in public health surveillance data. The years leading up to 2008 show an increasing trend, with the number of cases rising steadily, culminating in a peak that year. This peak, approaching the 100,000 marks, is markedly distinct from the surrounding data points, suggesting a significant increase in drug abuse incidents or a change in reporting practices during this period. Post-2008, the series exhibits a pronounced downward trend, indicating a substantial reduction in the number of reported cases. This reversal in trend is particularly noteworthy as it does not conform to the previous years' patterns, which displayed more gradual fluctuations. The sharp increase followed by a decrease suggests that external factors may have exerted a strong influence on drug abuse statistics during this time.

To understand the drastic decrease in the number of cases, Prophet was used for time series forecasting. Prophet was preferred over established models such as ARIMA and LSTM. ARIMA needs the data to be stationary, which requires some transformations on data necessary and is not as flexible as Prophet which considers real-life scenarios such as holiday effects into consideration. On the other hand, LSTM was not used as neural networks need a large amount of data for proper training.

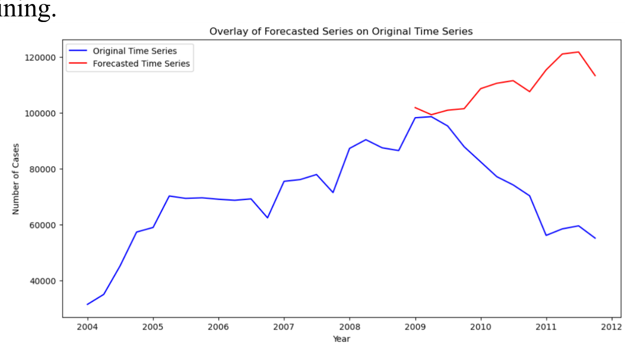

Fig. 7 Overlay of Predicted Time Series on Original Time Series

In examining the forecasted outcomes of drug abuse cases generated by the Prophet forecasting library against the actual observed data from 2004 to 2012, a significant divergence between predicted and actual values is evident. The forecast, depicted in red, extends from the culmination of the historical data in 2008 and illustrates a projection of continued growth in drug abuse cases. This projection contrasts sharply with the actual recorded data, which exhibits a marked decrease in cases post-2008. The Root Mean Squared Error (RMSE) of the forecast stands at approximately 40,368, and the Mean Absolute Percentage Error (MAPE) comes out to be 52.323%.

C. Granger Causality

An examination of the overlaid time series data, represented by two distinct yet seemingly congruent lines, depicts the trends of drug abuse cases and housing price indices over a selected period. The blue line illustrates the trend of drug abuse cases, while the red line corresponds to the housing price index. The visual synchronization of these lines suggests a parallelism in their trends, with both exhibiting similar patterns of rise and fall over time. This visual correlation prompts a hypothesis that movements in the housing market may reflect or even influence public health issues such as drug abuse.

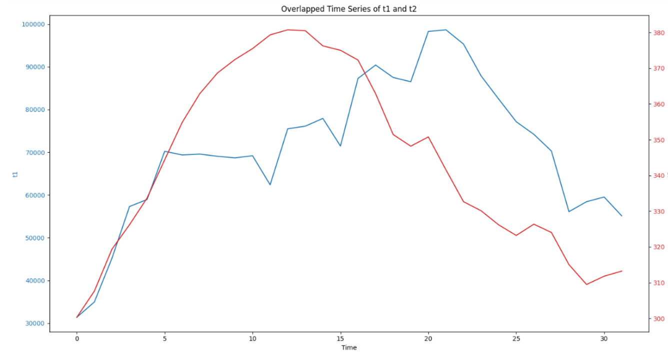

Fig. 8 Overlay Time Series of US Housing index on Drug Abuse

To statistically evaluate the potential predictive relationship between the two variables, a series of Granger causality tests were conducted. The Granger causality tests aim to determine whether one time series can forecast another by analyzing the lagged values of one series as predictors for the subsequent values of another.

The paper “The Impact of the 2008 Economic Crisis on Substance Use Patterns in the Countries of the European Union” showed that there was a downward trend in the substance use in European countries following the 2008 financial crisis. Based on this, it can be hypothesized that the same can be proven true for USA as well, as USA and European countries have similar cultures and demographics.[4]

D. Pattern Mining

The dataset, comprising substance abuse records from metro areas spanning the years 2004 to 2011, was systematically categorized. Each record was segregated by the metro area, and within each metro, by the substance of abuse. This organization facilitated the application of pattern mining algorithms tailored to each metro area over the specified period.

To discover prevalent patterns within the data, the Frequent Pattern Growth (FP-Growth) algorithm was employed. Unlike the Apriori algorithm, FP-Growth is adept at managing large datasets by constructing a compact representation known as the FP-Tree. This method significantly reduces the need for iterative database scans and candidate generation, rendering it particularly effective for the dataset at hand, which comprises over 2.2 million records. Upon application of the FP-Growth algorithm, frequent patterns associated with Alcohol and Drug abuse cases were extracted for each city within the eight-year span. These patterns were then subjected to a confidence analysis to ascertain their reliability. A threshold of 65% confidence was set to classify a pattern as reliable. Patterns not meeting this criterion were deemed unique.

Set operations were conducted to delineate patterns ubiquitously present across all metro areas throughout the duration of the study, which were termed 'interesting patterns.' Furthermore, patterns manifesting in at least nine metro areas across all years were identified and labeled as 'common patterns.' After the extraction of common patterns, the dataset was reevaluated to unearth new frequent patterns, now devoid of the common elements. A Jaccard similarity assessment was then implemented to quantify the resemblance between metro areas with respect to the newly identified frequent patterns.

E. *Classifier Comparision*

The dataset, spanning from 2004 to 2011, was curated to include categorical variables such as 'METRO', 'AGECAT', 'SEX', 'RACE', 'CASETYPE', 'DAYPART', and 'QUARTER'. The data was subsequently partitioned into training, validation, and test sets with a distribution ratio of 60:20:20, respectively.

Three classifiers were employed for the analysis: RandomForest, Naive Bayes, and CatBoost. RandomForest was selected for its proficiency in managing complex relationships within categorical data and its feature ranking capability. Naive Bayes was utilized for its assumption of feature independence, operational efficiency, and effectiveness with discrete variables. CatBoost was chosen due to its native support for categorical data, negating the need for extensive preprocessing and its enhanced model robustness via ordered boosting.

For model evaluation, metrics such as Accuracy, Confusion Matrix, Precision, Recall, and F1 Score were computed to facilitate a comprehensive comparison of the classifiers' performance.

F. *Feature Importance*

The study utilized the RandomForest classifier's feature importance functionality to identify the most critical predictive factors among 'METRO', 'AGECAT', 'SEX', 'RACE', 'CASETYPE', 'DAYPART', and 'QUARTER'. The analysis was executed in two phases.

Initially, an annual feature importance analysis was conducted for each metro area from 2004 to 2011. This allowed for the observation of temporal changes and the identification of trends in the significance of various features over time, providing insights into the evolving impact of factors such as age, gender, and race. Subsequently, the data from all years were consolidated for a cumulative feature importance analysis.

This comprehensive examination aimed to ascertain the overall importance of each feature throughout the study period, offering a holistic view of the most influential factors in substance abuse cases across different metropolitan areas. This methodical approach yielded a nuanced understanding of the key elements vital for predicting substance abuse patterns.

V. Results and Conclusion

A. *Time Series Forecasting*

The Root Mean Squared Error (RMSE) of the forecast stands at approximately 40,368, juxtaposed with the test dataset's mean value of 76,120.5. Additionally, the Mean Absolute Percentage Error (MAPE) is reported as 52.323%, which collectively signify a considerable deviation from the expected forecast accuracy. These metrics underscore the unforeseen nature of the drop in drug abuse cases, which was not anticipated by the model. The substantial size of these errors points to the unpredictability of the sudden decline rather than a deficiency in the forecasting method itself.

B. *Granger Causality*

The results of these tests across different lag periods exhibit a progression from non-significant to significant p-values, with a noticeable shift in causality strength as the number of lags increases. At lag period of four, the test statistics indicates significant p-value, suggesting that the past values of one series contain information that is useful in predicting the future values of the other series.

```
Granger Causality
number of lags (no zero) 1
ssr based F test:         F=3.1376  , p=0.0874  , df_denom=28, df_num=1
ssr based chi2 test:   chi2=3.4738  , p=0.0623  , df=1
likelihood ratio test: chi2=3.2926  , p=0.0696  , df=1
parameter F test:         F=3.1376  , p=0.0874  , df_denom=28, df_num=1

Granger Causality
number of lags (no zero) 2
ssr based F test:         F=1.1538  , p=0.3317  , df_denom=25, df_num=2
ssr based chi2 test:   chi2=2.7690  , p=0.2504  , df=2
likelihood ratio test: chi2=2.6486  , p=0.2660  , df=2
parameter F test:         F=1.1538  , p=0.3317  , df_denom=25, df_num=2

Granger Causality
number of lags (no zero) 3
ssr based F test:         F=3.0332  , p=0.0508  , df_denom=22, df_num=3
ssr based chi2 test:   chi2=11.9951 , p=0.0074  , df=3
likelihood ratio test: chi2=10.0386 , p=0.0182  , df=3
parameter F test:         F=3.0332  , p=0.0508  , df_denom=22, df_num=3

Granger Causality
number of lags (no zero) 4
ssr based F test:         F=2.1626  , p=0.1125  , df_denom=19, df_num=4
ssr based chi2 test:   chi2=12.7482 , p=0.0126  , df=4
likelihood ratio test: chi2=10.5058 , p=0.0327  , df=4
parameter F test:         F=2.1626  , p=0.1125  , df_denom=19, df_num=4
```

Fig. 9 Granger Causality Test Results

The significance of the Granger causality test at higher lag periods suggests that there is a statistical relationship between the housing price index and drug abuse cases, indicating that changes in one could be used to forecast changes in the other. This finding is particularly notable as it implies that economic conditions, as reflected by housing prices, could potentially serve as indicators for public health issues. The parallel trends observed in the overlaid time series analysis, supported by the Granger causality test results, provide preliminary evidence of a relationship between the housing price index and drug abuse cases. The statistical significance at higher lag values opens up avenues for further research to explore the underlying mechanisms of this relationship. This study underscores the importance of considering economic indicators when analyzing public health trends and points to the potential of using economic data as a predictive tool for public health outcomes.

C. *Pattern Mining*

a. *Interesting Patterns:*

i. Alcohol = ['alcohol only (age < 21)'], ['Young', 'alcohol only (age < 21)']

ii. Drugs = ['other'], ['adverse reaction'], ['Middle Age']

For alcohol-related cases, two notable patterns emerged: the prevalence of 'alcohol only' cases predominantly among individuals under the age of 21 and a significant occurrence of 'alcohol only' cases specifically among the 'Young' demographic within this age group. This highlights a critical issue of underage drinking, especially among younger individuals, indicating the need for targeted educational and prevention programs in this age group.

In the realm of drug abuse, the occurrence of 'adverse reactions' as a pattern point towards incidents that likely resulted from unintended or harmful effects of drug use. Additionally, the 'Middle Age' pattern implies a noteworthy prevalence of drug use in the middle-aged population, suggesting a demographic that may require increased attention in drug abuse interventions.

b. *Jaccard Similarity Heat Map:*

i. Jaccard Similarity for Alcohol Abuse Across States:

Fig. 10 Jaccard Similarity for Alcohol Abuse Across States

The Jaccard Similarity Heatmap for Alcohol provides a visual representation of the similarity in alcohol abuse patterns across various metropolitan areas. The heatmap indicates a range of similarity values from 0 to 1, where 1 denotes identical patterns and values closer to 0 indicate less similarity. Metropolitan areas such as Arizona (AZ) and Florida (FL) show high self-similarity, suggestive of comparable alcohol abuse characteristics in these regions. In contrast, there are metro areas with low similarity scores, highlighting distinct alcohol abuse patterns unique to those regions. This divergence suggests that interventions may need to be tailored to the specific substance abuse profiles of each metropolitan area.

ii. Jaccard Similarity for Drug Abuse Across States:

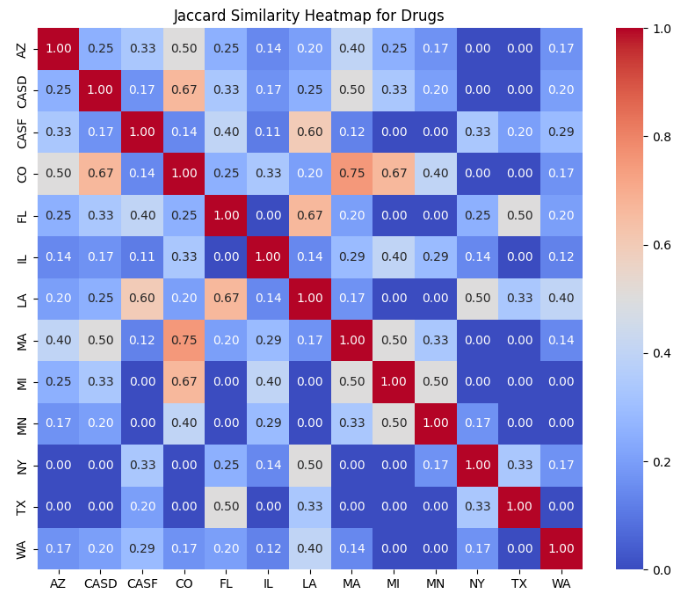

Fig. 11 Jaccard Similarity for Drugs Abuse Across States

The Jaccard Similarity Heatmap for Drugs provides a visual representation of the similarity in Drug abuse patterns across various metropolitan areas. There are many metro areas with low similarity scores, highlighting distinct drug abuse patterns unique to those regions. This divergence suggests that interventions may need to be tailored to the specific substance abuse profiles of each metropolitan area.

D. *Classifier Comparision*

```
Random Forest Classifier:

Confusion Matrix:
[[ 16018     13    130      0]
 [     1  11285  66695      0]
 [     4   8557 345178      0]
 [     0      0      0   6606]]

Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.99      1.00     16161
           1       0.57      0.14      0.23     77981
           2       0.84      0.98      0.90    353739
           3       1.00      1.00      1.00      6606

    accuracy                           0.83    454487
   macro avg       0.85      0.78      0.78    454487
weighted avg       0.80      0.83      0.79    454487


Accuracy Score:
0.8340986650883303
```

Fig. 12 Random Forest Accuracy Measures

```
Naive Bayes Classifier:

Confusion Matrix:
[[ 16018    133     10      0]
 [     0  56742  21239      0]
 [     0 111495 242244      0]
 [     0      0      0   6606]]

Classification Report:
              precision    recall  f1-score   support

     Alcohol       1.00      0.99      1.00     16161
Alcohol_DRUGS       0.34      0.73      0.46     77981
       DRUGS       0.92      0.68      0.78    353739
        None       1.00      1.00      1.00      6606

    accuracy                           0.71    454487
   macro avg       0.81      0.85      0.81    454487
weighted avg       0.82      0.71      0.74    454487


Accuracy Score:
0.7076330016040063
```

Fig. 13 Naïve Bayes Accuracy Measures

```
Catboost Classifier:

Confusion Matrix:
[[ 16018      7    136      0]
 [     0   9217  68764      0]
 [     0   5328 348411      0]
 [     0      0      0   6606]]

Classification Report:
              precision    recall  f1-score   support

     Alcohol       1.00      0.99      1.00     16161
Alcohol_DRUGS       0.63      0.12      0.20     77981
       DRUGS       0.83      0.98      0.90    353739
        None       1.00      1.00      1.00      6606

    accuracy                           0.84    454487
   macro avg       0.87      0.77      0.77    454487
weighted avg       0.81      0.84      0.79    454487


Accuracy Score:
0.8366619947325226
```

Fig. 14 Catboost Accuracy Measures

i. Random Forest Classifier:

High precision and recall for classes 0 and 3, indicating that it performs well on these classes. The F1-score for class 1 is very low, which implies that this model struggles with class 1 predictions. Overall accuracy is 0.834, which is quite high.

ii. Naïve Bayes Classifier:

Precision for class 1 (Alcohol_DRUGS) is notably low, indicating many false positives. Recall for class 1 is better than Random Forest, suggesting that it can identify more true positives for this class, but with less precision. Overall accuracy is 0.707, which is the lowest among the three models.

iii. Catboost Classifier:

Shows improved precision for class 1 compared to Naive Bayes, while maintaining high precision and recall for classes 0 and 3. F1-scores are generally high, except for class 1, but it is still better than the other two models for this class. Overall accuracy is 0.866, which is the highest among the three models.

The Catboost Classifier appears to be the best overall model. It has the highest overall accuracy and the most balanced performance across the different classes, indicating it is the most reliable model for this classification task. While the Random Forest Classifier also has high accuracy, its F1-score for class 1 is significantly lower than Catboost. The Naive Bayes Classifier lags in performance metrics compared to the other two models.

E. *Feature Importance*

i. Feature Importance for 2011:

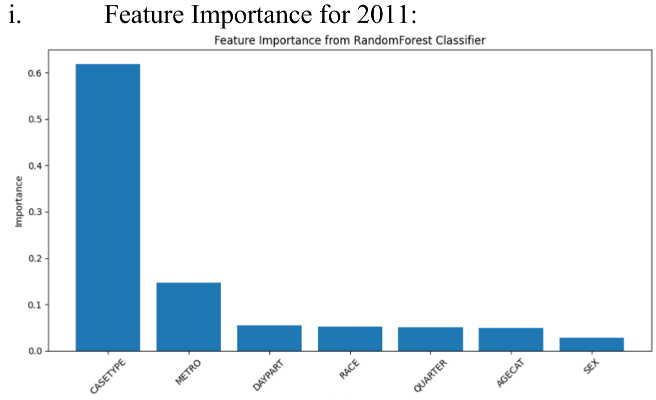

Fig. 15 Bar Plot of feature Importance for 2011

The feature importance remains relatively consistent across each year within the timeframe of 2004 to 2011.

ii. Feature importance on cumulative data

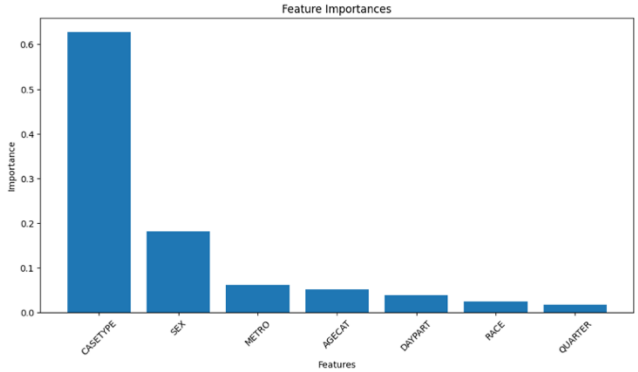

Fig. 15 Bar Plot of Feature Importance for cumulative data

The observed variations in feature importance between the two datasets underscore a crucial aspect of data analysis: features that may seem less significant in a short-term analysis can reveal greater importance when viewed across a longer time frame. The 2011 data suggest a scenario where 'CASETYPE' vastly overshadows other features, while the eight-year data from 2004-2011 reveal 'SEX' as a significantly more influential factor than it appeared in the single-year snapshot. This illustrates that features with limited impact on a yearly basis can, in fact, be critical when considering a broader temporal context. Such findings reinforce the concept that locally important features may not always hold the same level of global importance. Recognizing this distinction is vital for researchers and policymakers, as it emphasizes the need for comprehensive analysis over extended periods to fully understand and effectively address the multifaceted nature of issues like drug abuse.

Acknowledgment

I extend my heartfelt gratitude to Professor Jiebo Luo for his invaluable guidance and support during our project. His expertise and insightful feedback have been instrumental in shaping our success, inspiring us to strive for excellence. We are fortunate to have had the privilege of working under his mentorship, and we appreciate his dedication and encouragement.

Data Description

The Drug Abuse Warning Network (DAWN), operated by the Substance Abuse and Mental Health Services Administration (SAMHSA) of the U.S. Department of Health and Human Services, serves as a key source of data for my research. DAWN functions as a comprehensive public health surveillance system, meticulously tracking drug-related visits to hospital emergency departments (EDs) across the nation. This system not only covers cases directly caused by drugs but also includes visits where drugs play a contributing role.[3]

DAWN's data collection methodology is robust, relying on a longitudinal probability sample of non-federal, short-stay, general surgical and medical hospitals throughout the United States, each equipped with a 24-hour ED. This system is instrumental in monitoring trends in drug misuse and abuse, spotting the emergence of new substances and drug combinations, evaluating health hazards associated with drug abuse, and gauging the impact of drug misuse and abuse on the national healthcare system.[3]

The data encompassed in DAWN includes a wide array of information for each ED visit, such as demographic details, the drugs involved (with the dataset covering up to 16 drugs from 2004 to 2008 and up to 22 drugs from 2009 to 2011), toxicology confirmations, routes of administration, the type of case, and the patient's disposition following the visit. The DAWN website facilitates access to both national and metropolitan emergency department data tables, as well as comprehensive reports, which are integral to the analysis in my research.[3]

References

[1] J. E. Kallan, "Drug Abuse-Related Mortality in the United States: Patterns and Correlates," The American Journal of Drug and Alcohol Abuse, vol. 24, no. 1, pp. 103-117, 1998. DOI: 10.3109/00952999809001701

[2] World Drug Report 2018, United Nations publication, Sales No. E.18.XI.9.

[3] United States Department of Health and Human Services. Substance Abuse and Mental Health Ser vices Administration. Office of Applied Studies,"Drug Abuse Warning Network (DAWN), 2004,"Interuniversity Consortium for Political and Social Research,2015. DOI: 10.3886/ICPSR33041.v3.

[4] D. Dom, G. Geert, J. Samochowiec, S. Evans-Lacko, K. Wahlbeck, G. Van Hal, and D. McDaid, "The Impact of the 2008 Economic Crisis on Substance Use Patterns in the Countries of the European Union," *International Journal of Environmental Research and Public Health*, vol. 13, no. 1, p. 122, 2016. [Online]. Available: https://doi.org/10.3390/ijerph13010122.