What leads to incompletion: identifying lack of efficient treatment programs at the state level

Manqi Cai
University of Pittsburgh
mac538@pitt.edu

Hanyu Liu
Cornell University
hal4024@med.cornell.edu

Zhongyu Liu
Carnegie Mellon University
zhongyu2@andrew.cmu.edu

Yiwen Wang
Carnegie Mellon University
yiwenwan@andrew.cmu.edu

Abstract

We present an analysis aimed at identifying factors that contribute to the failure of treating substance abuse at the state level in the United states. Via exploratory data analysis, we uncovered patterns between variables and treatment outcome for different states. Furthermore, we employed machine learning techniques, including support vector machines (SVMs) and neural networks (NNs), to perform classification tasks. The SVM models demonstrate the influence of different features on treatment outcomes through the magnitude of the linear coefficients, while the NN model captures complicated non-linear relationships that can be visualized by Shapley values. The selected features were used to identify states that failed to treat specific types of patients. We conclude that type of substance and type of treatment service are the most important variable in determining the outcome of treatment, while states are failing in distinctly different types of treatment. States such as Kentucky, North Carolina and Hawaii are faced with more serious challenges.

1 Introduction

1.1 Problem statement

The main purpose of this project is to identify important factors that contribute to a given state's failure of treating substance abuse. The substance type, demographics, and socioeconomic status related to each substance abuse case all add to the challenge of facilitating efficient programs. For example, different states usually have different prevalent substances, which also require different strategies considering the fact that medication-based treatment programs are not available for all substances,

Since the numbers of cases and specialized treatment programs vary drastically among states, a detailed approach is needed to differentiate features that impact one state's success from those that benefit another. In this project, we applied multiple statistical inference and machine learning methods to do so, in order to identify states that are neglecting or having trouble with specific types of treatment.

1.2 Treatment records

The reason of discharge was selected to be our target variable. All records without a specified reason (marked as Other) were dropped. Records marked as treatment completed were classified as successful, and all other records were classified as failed. There are 3,886,871 records left in total.[1]

Other columns in the treatment records dataset were used as features. They can be further grouped into the following: demographics, marriage/employment/housing status, substance type and frequency, treatment details, etc. These are expected to be the core features in prediction of treatment completion.

1.3 Additional datasets

In addition to the patient's records, we also used the additional datasets as a backdrop to provide some information of the state (STFIPS) and statistical area (CBSA) assigned to each treatment case.

ACS [5] To ensure coverage for as many geographic areas as possible, we used the American Community Survey (ACS) 5-year datasets. They offer detailed estimates of demographics, housing situations and economics attributes of the population. These features constitute a broad picture of the population, signaling differences in income levels, education attainment and race/ethnicity.

Feature engineering We generated a state-based ACS dataset by combining estimates from all ACS sets grouped by state and year, where missing entries were filled by the average of other states' estimates in that year. For the CBSA-based dataset, we first combined estimates from all ACS sets grouped by county and year, filling missing entries with the average of other counties' estimates in that year. For each CBSA, estimates were then summed over all corresponding counties.

Facilities The treatment facilities dataset lists number of various treatment programs and specialized services in facilities by state, including features such as ownership, availablity of language assitance, etc.

Feature engineering All missing entries were filled with zero assuming non-availability. Features were then grouped by state and year and summed.

2 Exploratory data analysis

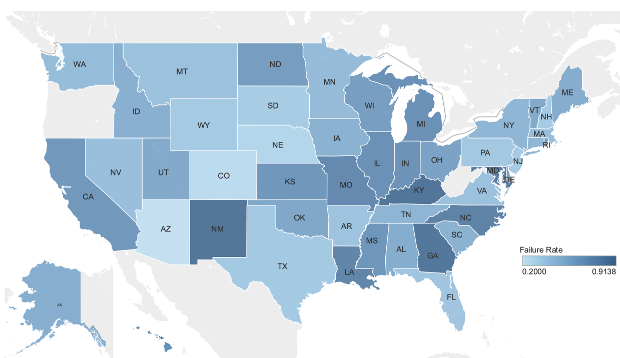


Figure 1: Geographic distribution of treatment failure rates for substance abuse in the US.

Prior to delving into machine learning algorithms, we conducted an extensive exploratory data analysis (EDA) of our dataset. This critical procedure allowed us to grasp the underlying patterns in our variables, unearth associations among different factors, and pinpoint variables that seem to contribute disproportionately to higher failure rates. Our examination spanned across critical indicators such as demographic profiles, substance type, and details of the treatment, exploring how each intertwined with our target variable - the reason for discharge. Moreover, we adopted a regional lens to compare these attributes, aiding our understanding of regional variances and nuances. For the sake of brevity, we have relocated some aspects of our exploratory data analysis, specifically those pertaining to demographic data, to the supplementary section.

The first step involved investigating the distribution of failure rates across different states in the US. Failure rate is defined as the proportion of cases not completed in all cases following Sec 1.2. From Figure 1, failure rates of substance abuse treatment in all states are displayed. Georgia (GA) and North Carolina (NC) have similar rates at approximately 82.5%, while New Mexico (NM) experiences a higher failure rate of 91.5%. In Kentucky (KY), the rate is the highest among the listed states at 93.6%. The rates in Louisiana (LA) and Maryland (MD) are approximately 80.9% and 85.2%, respectively. These figures underline the significant variation in treatment outcomes across different states, suggesting the need for location-specific strategies and interventions to enhance substance abuse treatment effectiveness. As depicted in Figure 2 below, there is substantial variation in the composition of primary substance use at admission across different U.S. states.

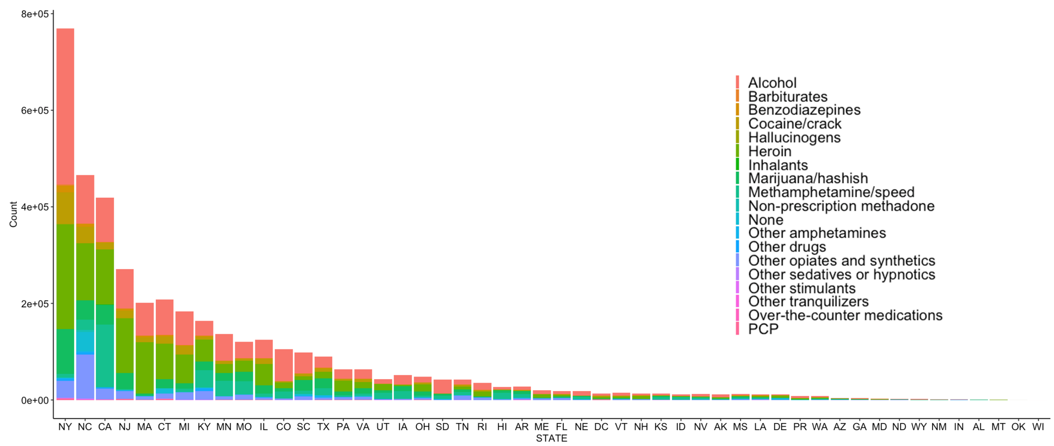


Figure 2: Distribution of primary substance use at admission across states in the US.

2.1 Effect size

In situations like ours where the number of samples considerably surpasses the number of features, conventional statistical tests may yield skeptical results d that may lead to misleading interpretations. Instead, to grasp the multifaceted influences of different variables on the failure rates across the country, we shifted our focus to evaluate the effect size, specifically the odds ratio between the successful and failed groups [4]. This serves as a measure of the relationship between the variables and the target. Differing from significance testing, effect size is less influenced by sample size and delivers a more accurate portrayal of the relationships within our data.

We calculated odds ratio (OR) between treatment completed and failure state by state under different variables. It was intriguing to note that about half of the states did not exhibit a significant with most of the variables when examined using OR (Figure 3A). This led us to hypothesize that in states boasting robust healthcare infrastructure, a patient's likelihood of completing treatment may not be predominantly determined by their individual attributes. Moreover, our analysis extended to identifying potential driving factors contributing to high failure rates. These would ideally be variables demonstrating a consistent significant difference in a majority of the states (Figure 3B). One example of such variables is the type of treatment service at admission and discharge (marked as SERVICES and SERVICES_D in our data).

3 ML methods

With the help of machine learning methods, we attempted to validate the available features' ability to predict successful treatment completion and establish relationships between features and treatment outcome that could be less obvious to the human eye at first glance.

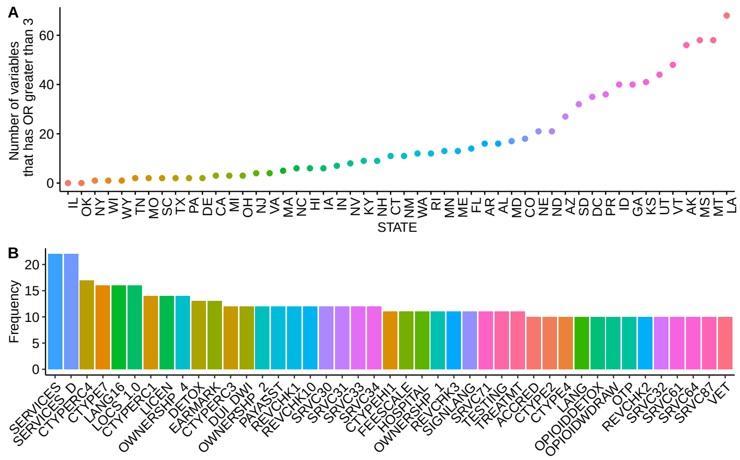


Figure 3: A): Scatter plot showing the distribution of states based on the number of variables that surpass an odds ratio of 3. B:) Bar plot depicting the frequency of top variables (frequency > 9) with an odds ratio greater than 3.

3.1 Classification

In classification, we dropped secondary and tertiary substance features to control the number of features and remove some redundancy. Substance flag features were kept. We also combined treatment records data with processed ACS and treatment facilities data. State and statistical area IDs were only used to join datasets, and excluded from classification.

A 75/25 split of data was used. The training set has 2,915,153 records, and the test set has 971,718 records. Both sets have a treatment completion rate of around 41.5%.

We tested penalized linear support vector machines (SVMs) and neural networks (NNs).

3.1.1 SVM classifier

Due to the size of the dataset, we used scikit-learn's implementation of SGDClassifier to train linear soft-margin SVMs. The training set was divided into batches of 10,000 data points, and before each epoch, ordering of batches was shuffled. Model training was terminated when training accuracy did not improve for a certain number of consecutive batches.

Using the hinge loss, the primal problem with penalty can be formulated as follows:

$$\min_{w,b} \frac{1}{2} w^T w + C \sum_{i}^{N} \max\left(0, 1 - y_i(w^T x_i + b)\right) + \text{penalty}$$

In the case of $\ell 1$ penalty, penalty $= \lambda |w|_1$, while penalty $= \lambda |w|_2^2$ in the case of $\ell 2$ penalty. SVM ($\ell 2$) showed better test performance (Table 3.1.2), as a result, it was selected as the baseline model.

3.1.2 Neural Network

To better capture the non-linear relationship between the features and target variable (i.e., treatment completion/failure), we built a feedforward neural network, including an input layer, four hidden layers, and an output layer. Specifically, the hidden layers have output sizes of (256, 512, 512, 64), where a dropout layer with a probability of 0.2 was added to the third hidden layer. The ReLU activation function is applied after each hidden layer to introduce non-linearity and allow the network to learn complex patterns.

Finally, the output layer takes the output from the fourth hidden layer (64 features) and produces the final output tensor. The sigmoid activation function is applied after the last linear layer to limit the output values to a range of 0 to 1, making it suitable for binary classification tasks.

As shown in Figure 4, the training process proceeded for a total of 100 epochs. Both training accuracy and test accuracy presented similar value during the training process, suggesting that the trained model has good generalization capability. The final prediction accuracy on the test set reached around 0.85 which is much higher than the linear SVM models' accuracies, indicating the non-linear patterns captured by the neural network significantly improves the prediction capability of the model. The neural network's F1 score and AUC score are both better than the baseline model, which shows that the NN's predictions are more balanced for both categories (treatment completion vs. failure).

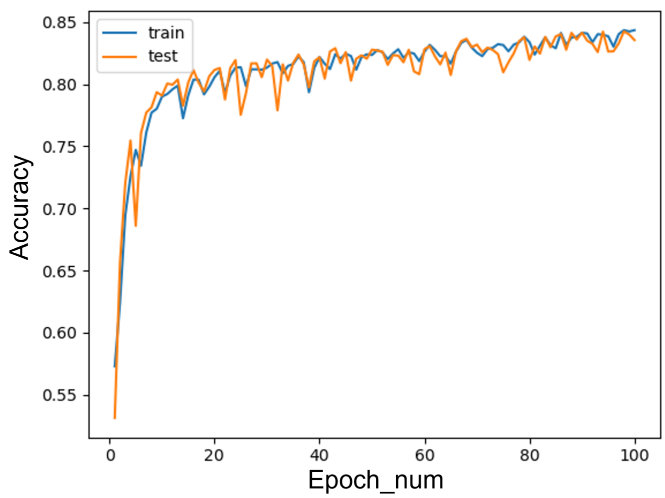


Figure 4: Accuracy during training process

Test results are listed in 3.1.2.

Table 1: Test results

Model	Accuracy	F1 Score	AUC
SVM ($\ell 1$)	0.758	0.709	0.751
SVM ($\ell 2$)	0.764	0.738	0.769
NN	0.848	0.787	0.815

3.2 Feature importance

Through examination of features' impact on predictions in each classifier model, we identified features that are important in generating the probability of treatment completion across the board. Methods of quantifying feature importance and results are discussed below.

3.2.1 Linear coefficients of SVMs

Even with regularization, the SVM classifiers still assigned large enough coefficients to most features. Therefore, we decided to compare the scale of linear coefficients to separate top features from the rest. For both SVM ($\ell 1$) and SVM ($\ell 2$), the top 10 features sorted by absolute value of corresponding coefficient are plotted below:

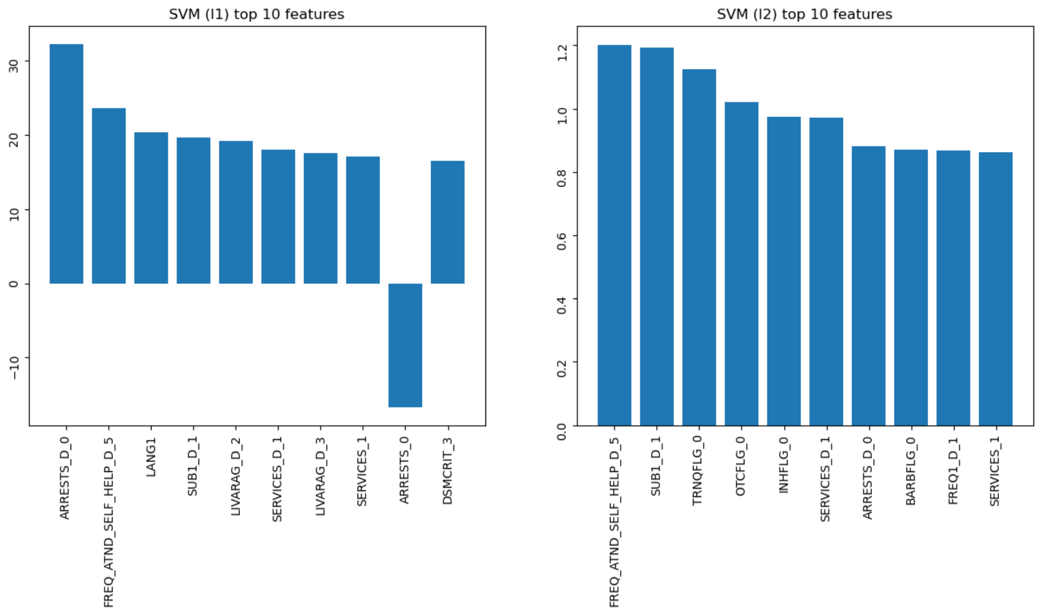


Figure 5: Top 10 features selected by SVMs.

Due to its linear nature, a trained SVM classifier defines a hyperplane (perpendicular to the support vector) that tries to separate successful cases from failed ones. It offers a more straightforward interpretation of features' negative/positive impact on the outcome, while ignoring complex interactions between features.

Some preliminary conclusions can be drawn from the selected top features and their linear relationships with the predicted probability. As expected, probability of a patient's treatment completion heavily depends on type of substance, treatment service and their criminal records. Patients that have no arrests 30 days prior to discharge, attend self-help groups, or do not have a primary substance at admission are more likely to complete their treatment programs. Detox patients and patients who did not use substance in the month prior to admission also tend to complete their treatment programs.

3.2.2 Shapley values of NN

Deep neural networks can be challenging to interpret due to their complex and non-linear nature. Here, we introduce Shapley value to understand how each feature affect the final output of the neural network. The Shapley value is a concept derived from cooperative game theory. It is used to assign a value to each feature or input in the neural network model, indicating its contribution to the prediction or output of the model [3]. If we let $\phi(X_i)$ denote the Shapley value assigned to feature i, the model prediction $f(X)$ can be written as

$$f(X) = \phi(X_1) + \phi(X_2) + \cdots + \phi(X_k)$$

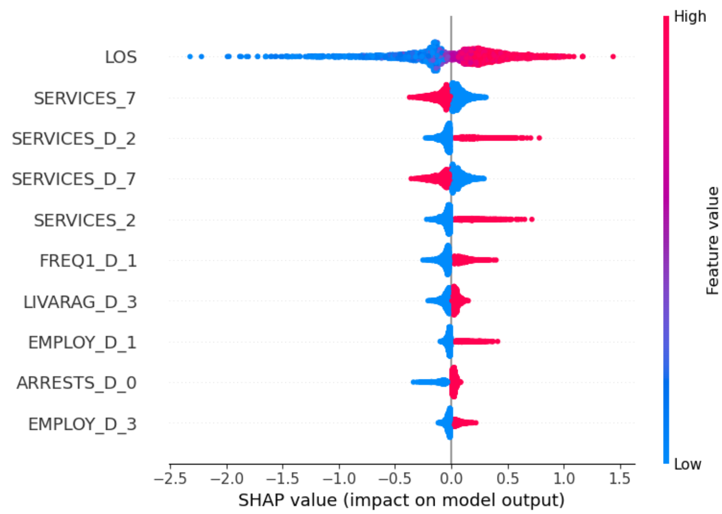


Figure 6: Top 10 contributed features in trained ANN model based on Shapley value. (All states, a positive SHAP value means a positive contribution to the complete of treatment program.)

Here, a positive SHAP value means a positive contribution to the predicted outcome of treatment program. In our model, the categorical features take 1 (True, red) and 0 (False, blue) and the continuous variable change from red to blue when it decrease from large value to small value (Figure 6).

As shown in Figure 6, length of stay (LOS) is determined as the most important feature for the trained ANN model and larger LOS value will lead to higher success rate of treatment. In other words, longer treatment session can result in better treatment results. Interestingly, receiving Service 2 (Detox, 24-hour, free-standing residential) seems to signal a higher probability of completion while receiving Service 7 (Ambulatory, non-intensive out patient) leads to a failure of the treatment program, suggesting further government support (i.e., financial and medical) is needed for further development of ambulatory care. Other main features suggest people who get employed have higher chances to complete the treatment program and people who get arrested tend to fail the program. What surprised us is that none of the demographic features were in the top contributors, potentially because personal attributes are much stronger indicators.

4 Identification of outstanding states

Based on classification analysis results discussed in 3.2, we selected variables from the treatment records dataset to divide cases into specific groups for each state. Each group's failure rate (proportion of cases not completed) in a given state was plotted against its proportion in all cases reported by that state to identify outsanding states that are failing specific types of treatment.

We considered the following variables: type of substance, substance use category (disordor/dependence/abuse), treatment service (detox/rehab/ambulatory), health insurance/payment (private/Medicaid/Medicare) and race/ethnicity. If a state has a high proportion/number of a given type of treatment case and also a high treatment failure rate for this group of patients, we identified it as an outstanding state. Selected plots are shown in 7.

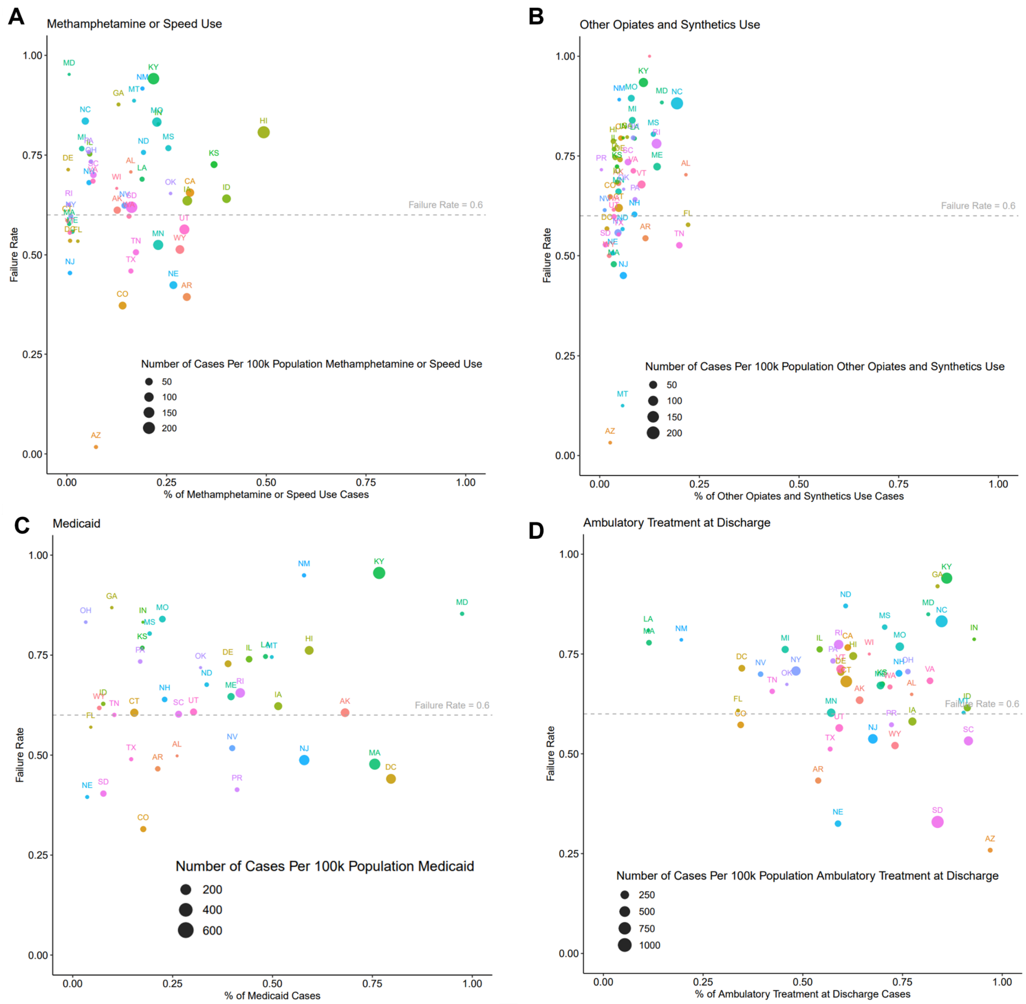


Figure 7: Failure rate vs. proportion of specified cases in all cases for each state. Average failure rate of all cases is around 60%, which is used as a threshold.

We identified Hawaii, North Carolina, Kentucky, and Idaho as outstanding states. Implications these statistical differences have will be discussed in Sec 5.

5 Conclusions

Ambulatory care is the one of the groups with high variation in percentage of cases (Fig 7 D). In general, higher percentage of ambulatory cases leads to a higher failure rate. States such as Kentucky and North Carolina are good examples.

In Hawaii, the number and ratio of methamphetamine cases are both alarmingly high (Fig 7 A), while its failure rate is also among the largest. This is supported by a recent study [2], which also proposed that economics, mental heath issues, and methamphetamine's relatively lower cost and higher availability all contribute to this issue.

Although the usage of opioids nationwide is on the rise, the proportion of opioid cases is still lower than other major substances such as alcohol and heroin. The state of North Carolina's failure rate and number of cases are both among the highest (Fig 7 B), signaling potential neglect of opioid treatment resources.

Cases where patients had Medicaid as the health insurance are also a group that clearly separates states (Fig 7 C). All with high percentages of Medicaid cases, Kentucky and Maryland had significantly larger failure rates compared to Massachusetts and D.C. This could be a result of less efficient healthcare systems and lack of additional financial support.

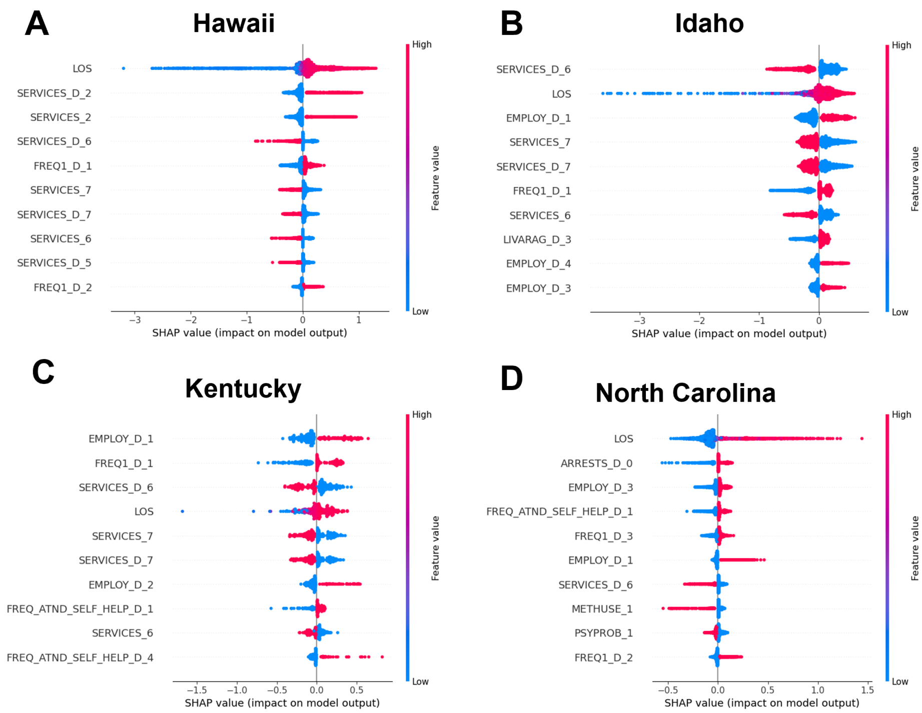


Figure 8: Top 10 contributed features in trained ANN model based on Shapley value for different states. (A positive SHAP value means a positive contribution to the complete of treatment program.)

For selected states, Shapley explanations of NN predictions are shown in Figure 8. For Hawaii, Idaho and Kentucky, receiving Service 6 (Ambulatory, intensive outpatient) or 7 (Ambulatory, non-intensive outpatient) tends to result in the failure of treatment, indicating the inability of these two types of service to provide efficient support for patients. Yet Service 6 is not identified in the national level (all states) analysis, suggesting this type of service may be influenced by substance abused locally. Methamphetamine takes a large portion in all three states' cases and it may be the reason why patients need Service 6.

Again, we found North Carolina to be an interesting case based on Fig 8 D, where methamphetamine only takes a small fraction of the total substance usage but it consistently causes failure of treatment. We argue that this sheds light on potential neglect of abuse of methamphetamine in this state. The wide abuse of heroin and other types of opiates in North Carolina could be the reason behind this lack of state-level attention.

In conclusion, to improve the quality substance abuse treatment, state governments and healthcare officials should focus on orchestrating substance-specific programs that cater to the needs of patients which also requires efforts to pinpoint predominant substances in the area. Other possible solutions include early intervention, medication assistance, ensuring accessibility of programs, decriminalization and controlling the cost of healthcare.

References

[1] S. Abuse and M. H. S. A. (SAMHSA). Treatment episode data set – discharges (teds-d) – concatenated, 2017 to present. `https://www.datafiles.samhsa.gov/dataset/teds-d-2019-ds0001-teds-d-2019-ds0001`, 2019. Accessed: yyyy-mm-dd.

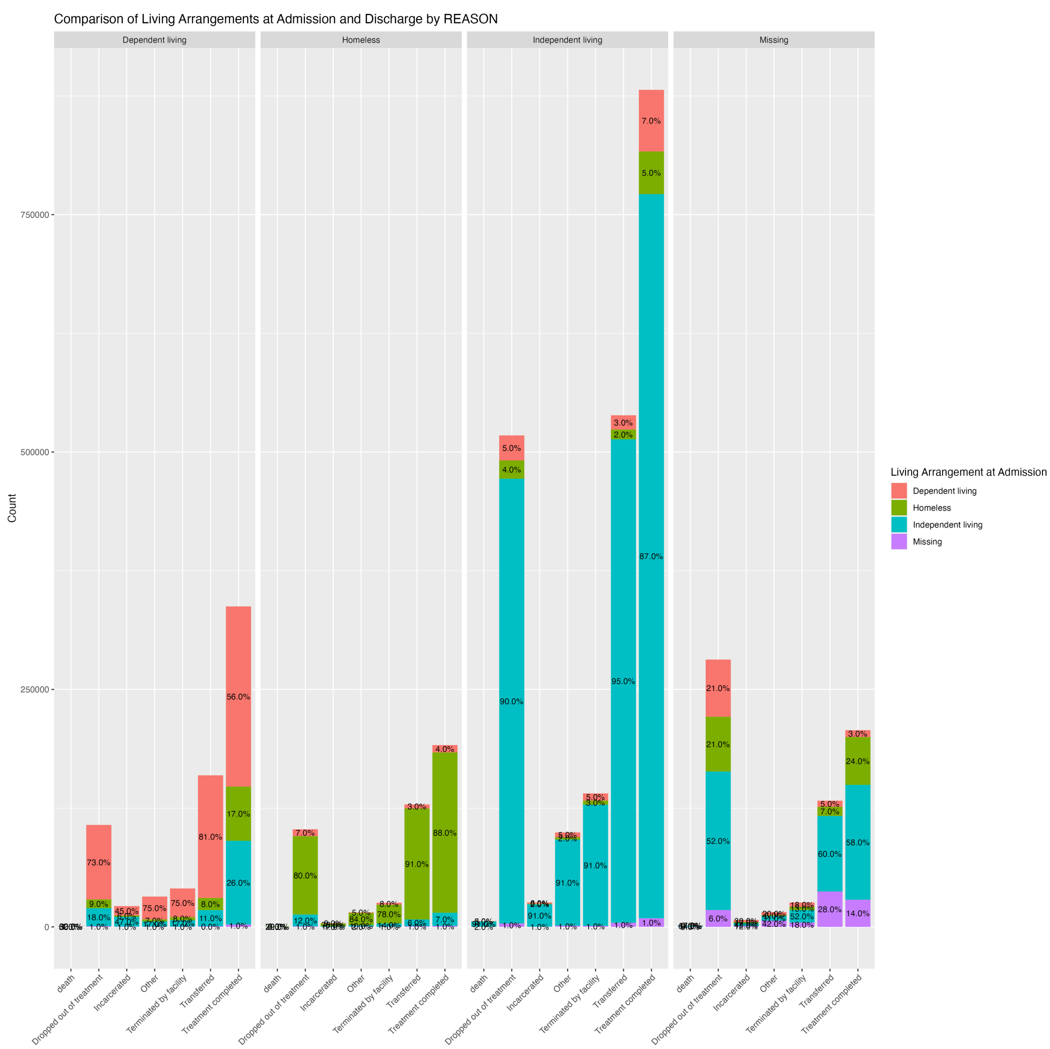
[2] M. Kiyokawa, M. Cape, and J. Streltzer. Insights in public health: Methamphetamine use during covid-19 in hawai ʻi. *Hawai'i Journal of Health & Social Welfare*, 80(5):117, 2021.

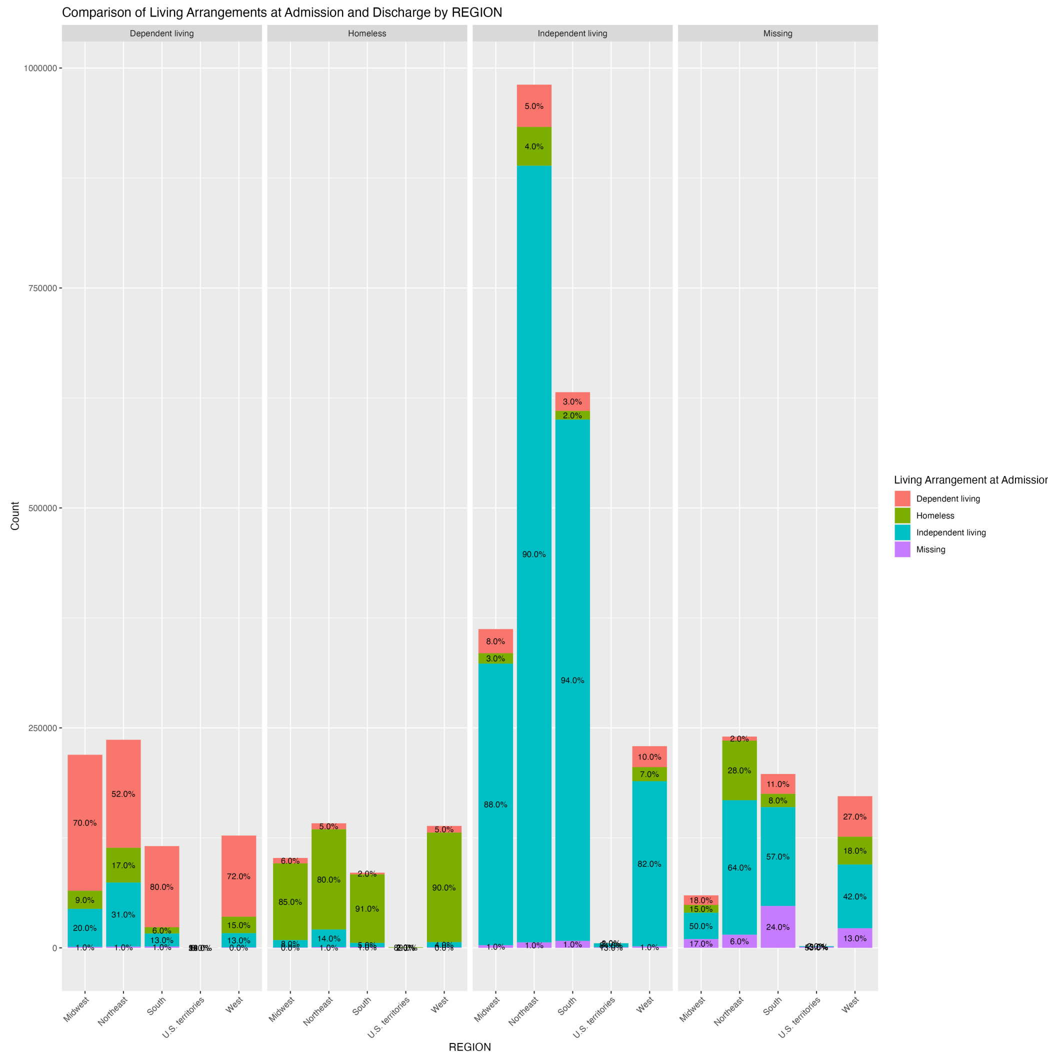
[3] S. M. Lundberg and S.-I. Lee. A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems*, pages 4768–4777, 2017.

[4] G. M. Sullivan and R. Feinn. Using effect size—or why the p value is not enough. *Journal of graduate medical education*, 4(3):279–282, 2012.

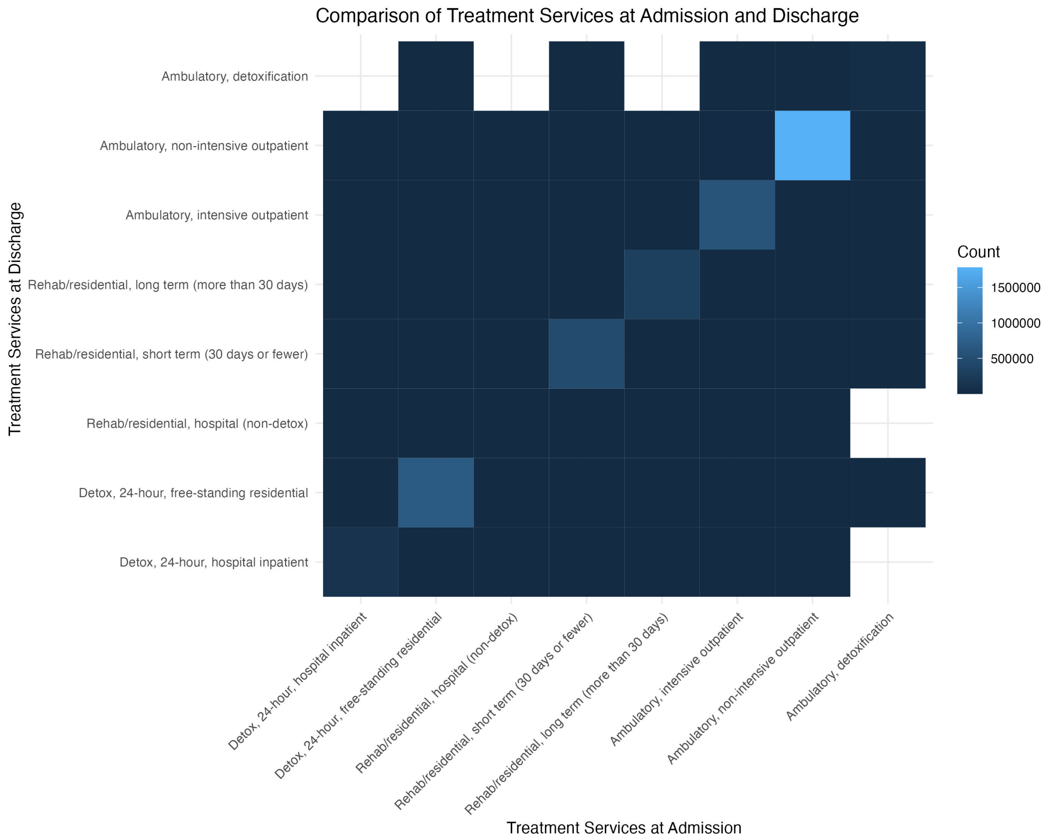
[5] U.S. Census Bureau. American Community Survey (ACS) Data. `https://www.census.gov/programs-surveys/acs/data.html`, 2023. Accessed: 2023-05-18.

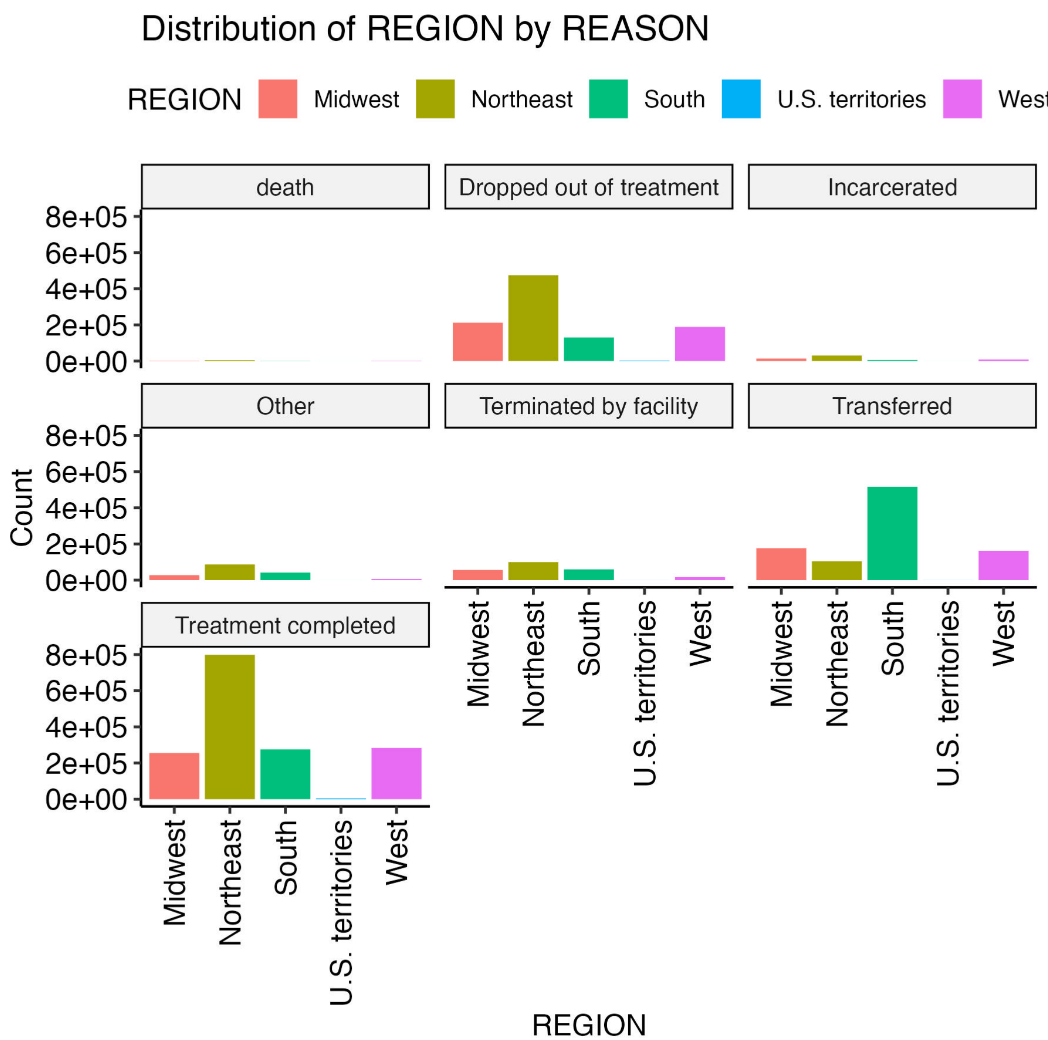
Supplementary material



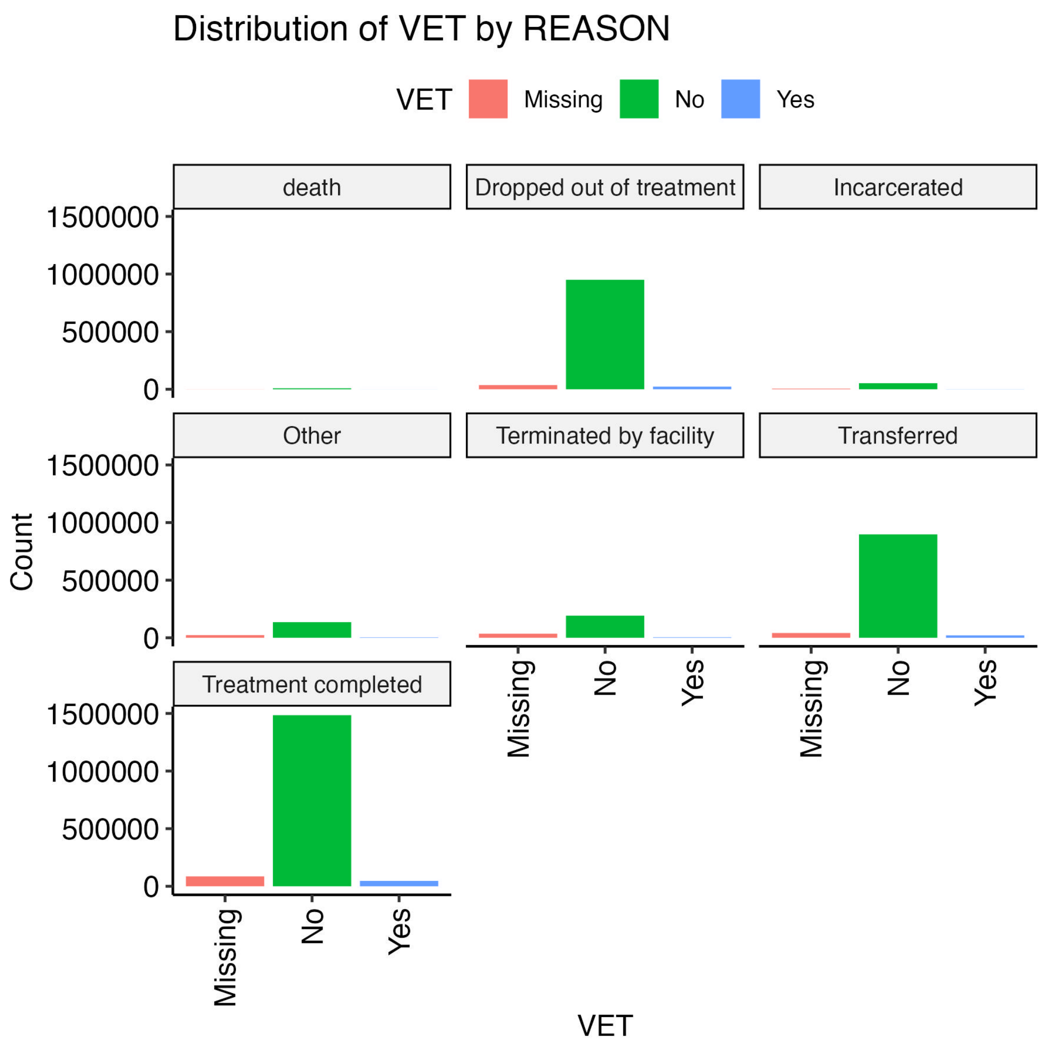


Comparison of Living Arrangements at Admission and Discharge by REGION
Dependent living
Homeless
Independent living
Missing
Count
REGION
Midwest
Northeast
South
U.S. territories
West
Living Arrangement at Admission
Dependent living
Homeless
Independent living
Missing


Comparison of Treatment Services at Admission and Discharge
Treatment Services at Discharge
Ambulatory, detoxification
Ambulatory, non-intensive outpatient
Ambulatory, intensive outpatient
Rehab/residential, long term (more than 30 days)
Rehab/residential, short term (30 days or fewer)
Rehab/residential, hospital (non-detox)
Detox, 24-hour, free-standing residential
Detox, 24-hour, hospital inpatient
Count
1500000
1000000
500000
Detox, 24-hour, hospital inpatient
Detox, 24-hour, free-standing residential
Rehab/residential, hospital (non-detox)
Rehab/residential, short term (30 days or fewer)
Rehab/residential, long term (more than 30 days)
Ambulatory, intensive outpatient
Ambulatory, non-intensive outpatient
Ambulatory, detoxification
Treatment Services at Admission



Distribution of REGION by REASON
REGION
Midwest
Northeast
South
U.S. territories
West
death
Dropped out of treatment
Incarcerated
Other
Terminated by facility
Transferred
Treatment completed
Count
8e+05
6e+05
4e+05
2e+05
0e+00
Midwest
Northeast
South
U.S. territories
West
REGION


Distribution of VET by REASON
VET
Missing
No
Yes
death
Dropped out of treatment
Incarcerated
Other
Terminated by facility
Transferred
Treatment completed
Count
1500000
1000000
500000
0
VET