Artificial intelligence to predict intimate partner violence perpetration

Chapter · November 2018

6 authors, including:

Robin Petering
Lens Co
72 PUBLICATIONS 837 CITATIONS

Mee Young Um
Arizona State University
21 PUBLICATIONS 200 CITATIONS

Nazanin Alipourfard
University of Southern California
17 PUBLICATIONS 269 CITATIONS

Rajni Kumari
Maharshi Dayanand University
19 PUBLICATIONS 19 CITATIONS

Artificial Intelligence to Predict Intimate Partner Violence Perpetration

Robin Petering, Mee Young Um, Nazanin Alipour Fard, Nazgol Tavabi, Rajni Kumari, Setareh Nasihati Gilani

Abstract

Violence is a complicated issue which has harmful physical and nonphysical consequences. In this article we are particularly investigating violence in intimate partner relationships among homeless youth. This phenomenon has many contributing factors thus it would be useful to know which of these variables have a more important role in determining how violent a person can be and also have a means to predict the chance of perpetration for a new person. In this article, we will find the most influential variables and also predict the chance of perpetration with the help of artificial intelligence. For this purpose, we used several supervised machine learning algorithms to build an intimate partner violence (IPV) perpetration triage tool which could be built and implemented in the field to identify young people who are at high-risk for engaging in violence perpetration.

1. Introduction

One-third of homeless youth (HY) experience some sort of physical intimate partner violence (IPV) and frequently young people experience bidirectional partner violence. One in four homeless youth have reported being both a victim and perpetrator in their current or most recent romantic relationship ([19]). Programs that address youth IPV frequently focus on victim and survivor support. Although supporting victims/survivors is necessary, offering only this type of support is not only paternalizing, with the underlying assumption that it is the victim's fault or responsibility to prevent IPV, it also does not effectively break the cycle of violence ([16]). IPV prevention programs should focus on reducing perpetration. Although batterer interventions exist ([5], [14]), they intervene after the incident of violence has occurred. These programs are typically court ordered, therefore the perpetrator participates in the intervention after there is an arrest or conviction. Additionally, batterer interventions have been shown to be ineffective and potentially do more harm than good in reducing IPV ([14], [18]). The current study proposes a method to predict the occurrence of adolescent and young adult IPV perpetration using Machine Learning (ML). Using data from homeless youth, an IPV perpetration triage tool will be built that can be implemented in the field to identify young people who are at high-risk for engaging in violence perpetration. This tool will encourage targeted services that will buffer the likelihood of engaging in violence such as mindfulness training to reduce impulsivity or general education in healthy conflict resolution skills. In order to solve the aforementioned problem different ML techniques will be used to predict the chance of perpetration for an individual and also most effective factors were identified using statistical approaches and were later compared with primary expectations from social work perspectives.

2. Problem Definition

Violence is a complex phenomenon that impacts adolescents and young adults across America. It occurs in multiple ways including interpersonal violence, intimate partner violence, gang and gun violence. Homeless youth experience all types of violence at higher rates than their housed counterparts ([20], [8], [3]). This is typically the result of many contributing factors including childhood experiences of trauma, subsistence survival strategies and exposure to perpetrators while living on the streets ([17], [23], [29]). The consequences of violence are severe, besides the proximal consequence of severe injury and or death, violence can also cause non physical ailments as posttraumatic stress disorder, depression and externalizing behavior, delinquency and aggression). It is clear that reducing violence in the lives of homeless youth is imperative and will contribute to a young person's ability to safely and successfully exit the streets and lead a long a productive life in society. However, public health and social interventions to reduce violence within adolescent and young adult populations are difficult ([21]), because, again, this phenomenon is complex with many intrinsic and extrinsic contributing factors.

The public health model recognizes three levels of prevention: primary, secondary, and tertiary ([33]). Primary prevention is designed to reduce incidence by preventing the first occurrence of an event. Secondary prevention is designed to decrease the prevalence of a problem after its onset and often includes interventions targeting populations at greatest risk of harm. Tertiary prevention occurs once a problem is clearly evident and causing harm ([12], [33]).

There are primary IPV prevention mechanisms which is curriculum based which focuses on teaching awareness about violence, promote healthful behavior, getting them involved in mindfulness exercises and teaching conflict resolution skills. But the IPV problem among homeless populations is more complex, given that the youth are not usually going to school and are not a regular part of any community to get access to those primary intervention techniques.

Substance abuse, including alcohol which is prevalent in homeless youth, is directly linked to IPV ([27]). Of all the reported cases of domestic violence in Australia in 2006, 51% of them were flagged as alcohol related ([13]), which might be even more because many times the report might not be flagged. In the homeless communities, substance abuse is rampant and play a key role in crime ([27]). These are the evident factors which can be easily seen as the cause of IPV; however, these can be prevented by using some manual curriculum based primary interventions. But there are many hidden factors which might be causing this behavior in perpetrators in relationships they are part of.

Even the primary intervention systems use simplistic and generic methods which might not be that effective in the long run. There are several community programs to help the IPV victims but not many programs are out there which are focussed on identifying and counseling the perpetrators and batterers. Often, perpetrators learn their violent behavior by witnessing or being exposed to domestic violence during their formative years. The personal background and upbringing of each perpetrator play a key role in their future violent behaviors. Hence, these might go unnoticed in traditional intervention techniques, which highlights the need for newer and efficient ways to intervene. Most of the intervention and prevention programs for IPV for homeless have limited funding and it is important that we use the allocated resources in the most efficient way. AI supported methods are expected to provide targeted interventions which are more effective.

Currently, there are several organizations that are working on many of the worlds hardest problems: combating child exploitation, disrupting illicit networks, delivering humanitarian aid in the wake of conflict and natural disasters, and more. When we talk about artificial intelligence here it spans the core artificial intelligence, machine learning and data mining using machine learning methods. AI and data mining helps revolutionize the way we use data in pursuit of helping the society. Further, unlike many other proposals to improve society, machine learning tools can be easily scaled to larger demographics depending on the requirements.

3. Dataset Description

As part of a longitudinal study of Los Angeles area homeless youth, drop-in service seeking youth completed a self-administered questionnaire. The presented results are from the third panel of data collection (sample size; N=452). The Revised Conflict Tactics Scale (CTS2) was used to assess physical IPV perpetration. The sample was limited to youth who answered the questions related to IPV (99 youth either did not answer the corresponding questions or were never in a relationship which narrows down the data to 353). The research team approached all youths who entered the service agencies during the data collection period and invited them to participate in the study. The selected agencies provided weekday services to eligible HY, including basic needs, medical and mental health services, case management, and referrals and connections to other programs such as housing services. Each youth signed a voluntary consent form and a consistent pair of research staff members was responsible for all recruitment to prevent youths from completing the survey multiple times during each data collection period per site. The questionnaire asks the respondent about their personal life, their interactions with other people, where and how they live, the quality of their relationship and sexual life if they have any partner, etc. These data are unique in that they include the Revised Conflict Tactics Scale ([1]), which was to assess physical IPV perpetration in each participants most recent intimate relationship. The CTS2 is the most widely used instrument in research on interpersonal violence and includes data on perpetration and victimization across various domains of violence (i.e. physical, emotional, relational, sexual and threatening). The current studys primary outcome variable was physical perpetration that includes a range of items from "I slapped my partner" (minor) to " I kicked my partner" (severe). The original dataset had over 1000 variables. The number of variables in comparison with the number of participants is too large for using Machine Learning techniques. Therefore, unrelated variables were removed from the dataset for analyses, which resulted in decreasing the number of potential predictor variables to 26. The variable name and description are presented in table 1.

4. Data Analysis

The data analysis occurred in several stages. First, we did an analysis based on what is found in previous social science literature on IPV perpetration. Second, we did a p-value and lasso technique analysis. Each stage in this analysis phase will be used to identify the most important features that could be included in a IPV perpetration triage tool.

Age	Age at the time of interview
Exchange Sex	Exchanged sex for money or other items
Children	Number of children
Suicide Attempt	Suicide attempt in previous 12 months
Homeless Age	Age of first homelessness
Jail	Ever been in Jail
Juggalo	Identifies as a Juggalo (fan of musical group ICP)
Male	Identifies as male
LGBQ	Sexual minority
White	Of caucasian race/ethnicity
Literal Homeless	Is literally homeless (sleeps on street, park, car, etc.)
Weapon	Carried a weapon in previous month
Violence	Engaged in interpersonal violence (physical fight) in previous 12 months
Gang	Is current or former gang member
PTSD	Has symptoms of PTSD
Depressed	Has symptoms of depression
Community Violence	Witnessed community violence during childhood
Sexual Abuse	Experienced childhood sex abuse
Physical Abuse	Experienced some form of physical child abuse (physical & witness family IPV)
Lonely 1	How often do you feel that you lack companionship?
Lonely 2	How often do you feel left out?
Lonely 3	How often do you feel isolated from others?
Hard Drug Use	Hard drug use in past 30 days
Foster Care	Ever in foster care
Job	Currently has a job
IPV perpetration	Is a violent relationship or not? (The outcome variable which we want to predict)

Table 1: Description of Dataset Variables

4.1. Theoretical quantitative analysis

Before performing machine learning analyses, we established a baseline analysis as a comparison. We identified variables in the current data set that correspond with risk factors and predictors that have been found in previous literature. Age, race and gender were included in the baseline analysis as statistical control variables. For adolescents and youth, IPV experiences are not predicted by gender which is contrary to adult experiences of IPV ([10]). The most common predictor for youth IPV perpetration is experiences of childhood maltreatment or abuse. A large body of literature has suggested that exposure to maltreatment in childhood is related to some form of IPV in later life ([2], [4], [7],[15], [31]). The link between maltreatment and IPV is often referred to as intergenerational transmission of violence ([32], [30]), which is derived from Banduras (1979) social learning theory. This framework suggests that children exposed to violence, either as a victim or witness, are more likely to use violence as a tactic for conflict resolution because this has been modeled in intimate relationships during childhood. A statistical analysis was performed including: Age, White, Male, Community Violence, Sexual Abuse, Physical Abuse.

4.2. Statistical Analysis

Multiple methods were used to uncover the most important features. P-value and lasso techniques are described here; AI algorithm techniques, such as SVM and random forest, are described in the Methodology section below.

4.2.1. P-value

The p-value is used in the context of null hypothesis testing in order to quantify the idea of statistical significance of evidence ([28]). We calculated this value for each feature to achieve a ranking of the most influential ones.

4.2.2. Lasso

A method which is useful in determining the most important features is LASSO([26]). LASSO is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the statistical model it produces. In other words, it will give us the variables that are most influential on the target variable. Since we have many variables, using this method will help us to reduce the number of features to consider in the final model and in our triage tool at the end since it is not feasible to consider all the features that we have now.

4.2.3. Support Vector Machines

SVM is a well known predictive model. Assuming features form the dimensions in space (In this way each data point would be a point in our space), SVM finds the hyperplane or a set of hyperplanes which best separates the data points of different classes. Best separator for a SVM model is the one with largest margin (distance of the hyperplane to the nearest data point of any class). Having the trained model, a previously unseen data point is mapped into the same space and based on its position relative to the hyperplane, it is marked as positive or negative (in a binary classification). For some datasets different classes are not linearly separable, to generate non-linear classifiers kernels are introduced. Different kernels and also different parameters for each kernel are among parameters we tune to get the best model. The best SVM model for our dataset is a rbf kernel. Though this model is clearly a predictive model, its other features lead us to try it to find the most effective features. The hyperplane constructed by SVM gives a weight to each feature in space. These weights or coefficients can be used to determine which features had a bigger impact in separating the positive and negative values and which features were irrelevant.[22]

4.2.4. Random Forest

Random forest is a popular algorithm for feature ranking. In the structure of decision trees, leaves represent class labels, or IPV_perpetration in our data, and branches represent conjunctions of features that lead to those class labels.

4.2.5. Final Rankings

In table 2, 7 most effective features based on each algorithm are given. Lasso and T-test are described in the previous subsections, SVM and Random Forests are described in methodology. The repetitive features in algorithms are used as important features. With this approach, the most important features are: PTSD, depressed, violence, hard_drug_use, lonely3, exchange_sex.

	SVM	**Random Forest**	**T-test**	**LASSO**
1	lonely2	**depressed**	**ptsd**	**violence**
2	**hard_drug_use**	**ptsd**	**depressed**	**ptsd**
3	**ptsd**	homeless_age	**hard_drug_use**	**depressed**
4	**exchange_sex**	Age	**violence**	**hard_drug_use**
5	juggalo	**violence**	**lonely3**	gang
6	**lonely3**	weapon	physical_abuse	**exchange_sex**
7	sexual_abuse	**hard_drug_use**	suicide_attempt	LGBQ

Table 2: Ranking of the features in the decreasing order of importance

After removing ptsd from variables set, we run the ranking algorithms again on the data. By removing ptsd, other variables can increase their power in the data. Based on the results, most important features are: depressed, hard_drug_use, violence, gang, physical_abuse, lonely 3.

	SVM	**Random Forest**	**T-test**	**LASSO**
1	**hard_drug_use**	**depressed**	**depressed**	**depressed**
2	exchang_sex	homeless_age	**hard_drug_use**	**hard_drug_use**
3	**gang**	**violence**	**violence**	**violence**
4	lonely2	Age	**lonely3**	**gang**
5	juggalo	**hard_drug_use**	**physical_abuse**	**physical_abuse**
6	**lonely3**	**lonely3**	suicide_attempt	LGBQ

Table 3: Ranking of the features, after excluding PTSD, in the decreasing order of importance

5. Methodology

In this section, we will describe the final features we used in our ML techniques, our baseline model which we will compare our results with, and a short description of all the Machine Learning algorithms that we used.

5.1. Features

As discussed in section 3, we have 26 features in our dataset. In the data analysis part, we were trying to find the most important variables that we have. Based on the results, we found out that ptsd has a major contribution in determining the target value. But in fact, this is a controversial issue because the directionality of this variable to perpetration is unclear. There is an immense research body that shows that PTSD is a direct consequence of violence ([25]) and without longitudinal data there is no way to confirm this is not what is happening. Therefore, PTSD was excluded from the predictive modeling. In order to see the importance of this variable and to determine if we can perform as good as we did when we are not considering this variable at all, we are defining three feature sets to explore our machine learning algorithms with. The first feature set includes all the features. The second one is just considering the most important ones based on the results of our data analysis. And the third feature set is using the most important features when we remove the ptsd from the initial feature set and then rank the other features based on their importance. Results of running our algorithms on these three distinct feature sets are reported in section 6.

5.2. Baseline

Due to lack of prior work in the domain of our problem we define literature based baseline. We used logistic regression algorithm for the following variables: Age, white, male, community_violence , sexual_abuse, physical_abuse.

5.3. Learning Algorithms

Predictive models use known results to develop (or train) a model that can be used to predict values for unseen data. Modeling provides results in the form of predictions that represent a probability of the target variable based on estimated significance from a set of input variables. Using the previously described features, we tried several approaches by using different machine learning algorithms to build our classification models. Some of the supervised machine learning system that we explored are: Logistic Regression, Support Vector Machines, Random Forest, Neural Network, Deep SVM. We trained and tested these models using K-fold (5 fold) cross validation and in order to evaluate the models we will use metrics such as precision, recall, F1-score, ROC AUC (Area Under Curve) and accuracy.

5.3.1. Neural Networks

Artificial neural networks is a computational model used in several computational techniques in computer science and other research disciplines. They are analogous to biological neurons where connections between neurons carry an activation signal of varying strength. If the sum of the incoming signals are strong enough, the neuron becomes activated and it is passed on to the next layer([24]). We chose neural networks for our model because it has an ability to generalize and respond to unexpected inputs and patterns. But at the same time neural nets usually need massive amounts of data to train which might be the reason we are not getting the best results. For our model we are using 10 layers and 10 hidden nodes.

The structure of a decision tree gives a good overview of important features, thus beside classification, we use random forest for feature ranking.

5.3.2. Logistic Regression

Logistic Regression is a linear classifier which finds the hyperplane which can best describe the data (by mapping the data points onto that hyperplane) and then separates the data points by a threshold on that hyperplane which again best separates the data points of different classes.([9])

5.3.3. Deep Support Vector Machine

The next model we used for classification is two class locally deep support vector machine. It is a supervised learning method and is available as a part of Microsoft Azure Toolkit. It creates a two class classification model and use non linear Support Vector Machine (SVM) optimized to achieve higher efficiency in cases of larger training sets([11]). Use of a local kernel function enables the model to learn arbitrary local feature embeddings, including high-dimensional, sparse, and computationally deep features that introduce non-linearities into the model. The kernel function that is used for mapping data points to feature space is designed to reduce the training time while maintaining the classification accuracy. LD-SVMs are best used when we have a complex data such that linear models (viz. Logistic regression) perform poorly and data is very complex.

6. Results

We performed our experiments with several algorithms. To standarize the experiments we used 10 fold cross validation of handling the test and training sets. The following metrics were used to determine which algorithms worked best for our experiments.

1. Accuracy - Accuracy is the proportion of true results to the total number of examined data.

2. Precision - Precision is the probability measure that a retrieved information is relevant.

3. Recall - Recall is the probability measure relevant information is retrieved in a search.

4. F1-score - A measure that combines precision and recall is the harmonic mean of precision and recall, the traditional F-measure or balanced F-score:

(2*(Precision*Recall)/(precision+recall)

5. ROC AUC - Area under the curve

Since we have imbalanced data the best measure for comparing different algorithms are

F1-measure and ROC AUC. As shown in the chart both F1 score and ROC AUC measure of the SVM classifier (with RBF kernel) have the highest value and it gives the best results so that is the model we propose for our problem.

	Deep SVM	**Neural Net**	**Random Forest**	**Logistic Regression**	**SVM(RBF kernel)**	**Baseline**
Accuracy	0.65	0.62	0.66	0.61	**0.66**	0.55
Precision	0.5	0.47	0.46	0.42	**0.48**	0.36
Recall	0.48	0.45	0.34	0.56	**0.6**	0.52
F1	0.48	0.44	0.44	0.48	**0.53**	0.42
ROC_AUC	0.64	0.6	0.64	0.62	**0.69**	0.54

Table 4: Results of prediction of classification algorithms for all variables (feature set 1)

For defining the most important factors, we find factors which are more common in our four rankings available in table 2 . The most important variables are ptsd, depressed, violence, hard_drug_use, lonely3, exchange_sex.

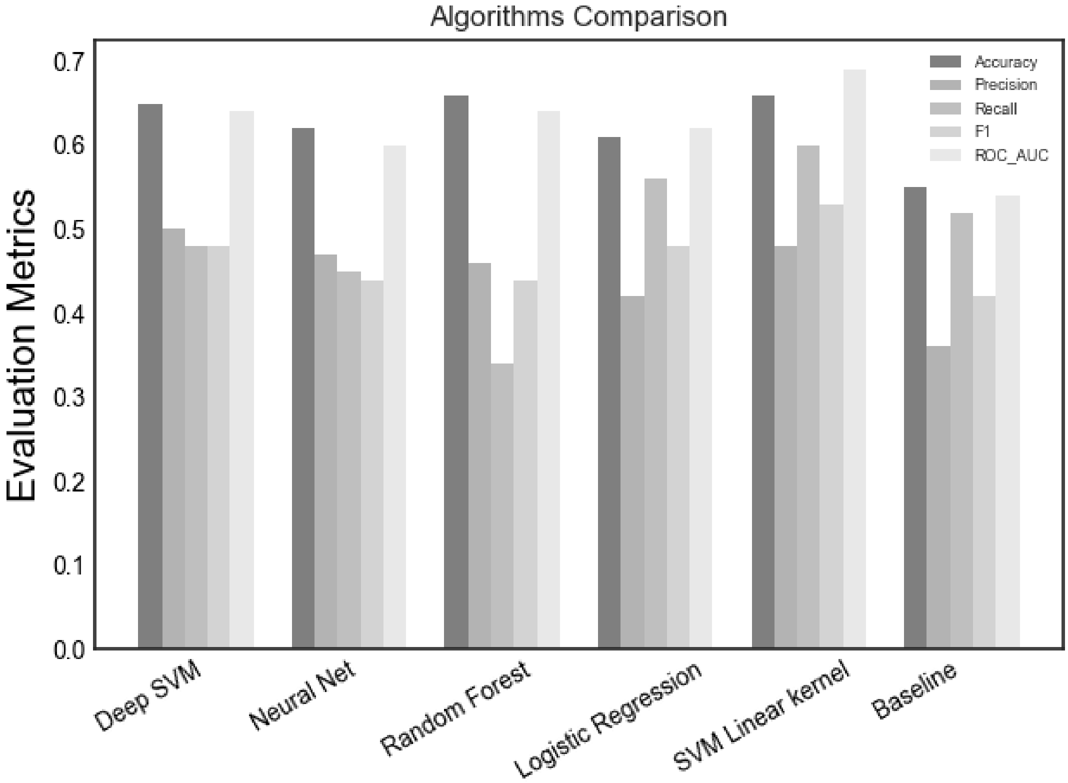

Figure 1: Comparison of different Algorithms for feature set 1

	Deep SVM	**Neural Net**	**Random Forest**	**Logistic Regression**	**SVM(RBF kernel)**	**Baseline**
Accuracy	0.69	0.71	0.62	**0.66**	**0.58**	0.55
Precision	0.49	0.6	0.41	**0.48**	**0.43**	0.36
Recall	0.36	0.42	0.43	**0.63**	**0.9**	0.52
F1	0.41	0.47	0.42	**0.54**	**0.58**	0.42
ROC_AUC	0.64	0.71	0.67	**0.69**	**0.67**	0.54

Table 5: Results of prediction of classification algorithms for important variables including PTSD(feature set 2)

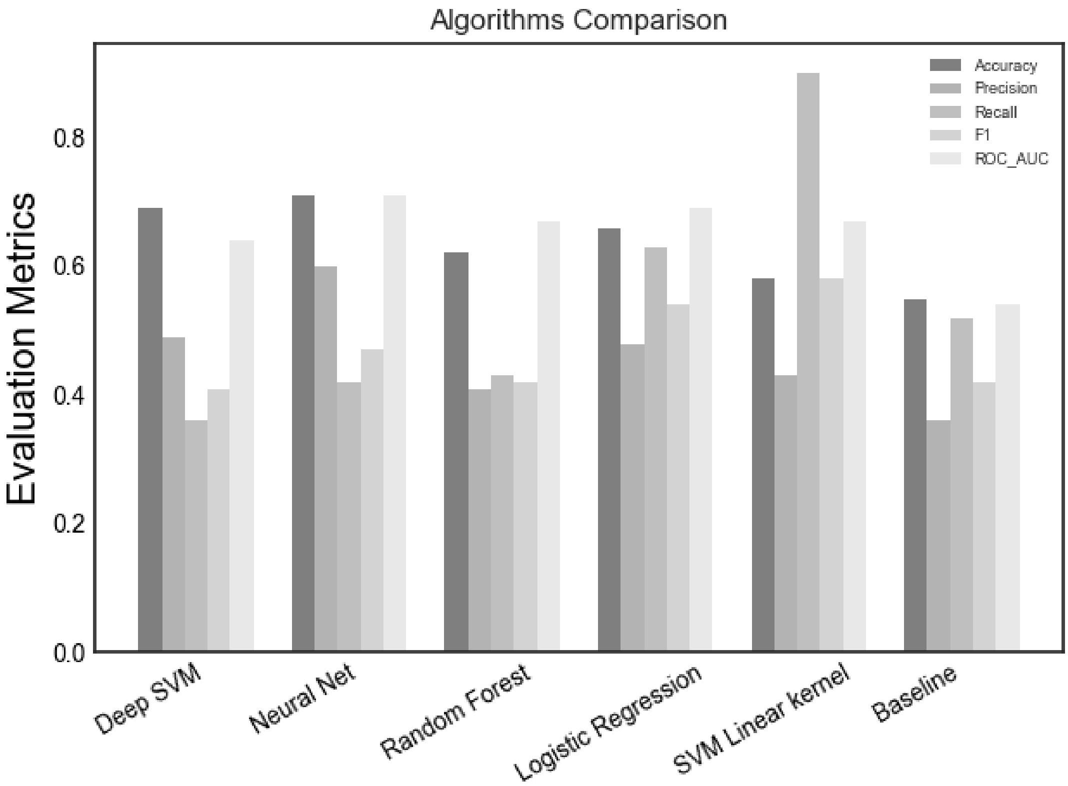

Figure 2: Comparison of different Algorithms for feature set 2

As mentioned before, after removing ptsd from features we then extracted the most important features from the rest of the features. Hence we got a new feature set: depressed, hard_drug_use, violence, gang, physical_abuse, lonely3. Table 6 shows the results using this feature set.

	Deep SVM	**Neural Net**	**Random Forest**	**Logistic Regression**	**SVM(Linear kernel)**	**Baseline**
Accuracy	0.66	0.66	0.61	0.63	**0.63**	0.55
Precision	0.43	0.5	0.41	0.45	**0.46**	0.36
Recall	0.26	0.36	0.41	0.58	**0.58**	0.52
F1	0.3	0.4	0.41	0.51	**0.51**	0.42
ROC_AUC	0.67	0.64	0.63	0.66	**0.67**	0.54

Table 6: Results of prediction of classification algorithms for important variables excluding PTSD(feature set 3)

To summarize the data presented in this three tables and figures, Table 4 and Figure 1, Table 5 and Figure 2, Table 6 and Figure 3, we observe that SVM gave us the best results in two of our feature sets. When using all the variables as our feature set, we got the highest ROC and F-1 score (ROC= .69, F1= .53) using SVM with RBF kernel. SVM with linear Kernel gave us the best ROC and F1 score when we used the significant variables (after removing PTSD from our initial variables). (ROC=.67 , F1= .51) When exploring with the significant variables including PTSD, we got the highest ROC using Logistic Regression method (ROC= .69). In this case, SVM with RBF Kernel gave us the highest F1 score of 0.58. In general, we didn't get great results in terms of F-1 measure and ROC curve. However, these might have resulted from our limited and sparse dataset. We had too many variables compared to the size of our dataset and also we had a lot of missing values for some variables.

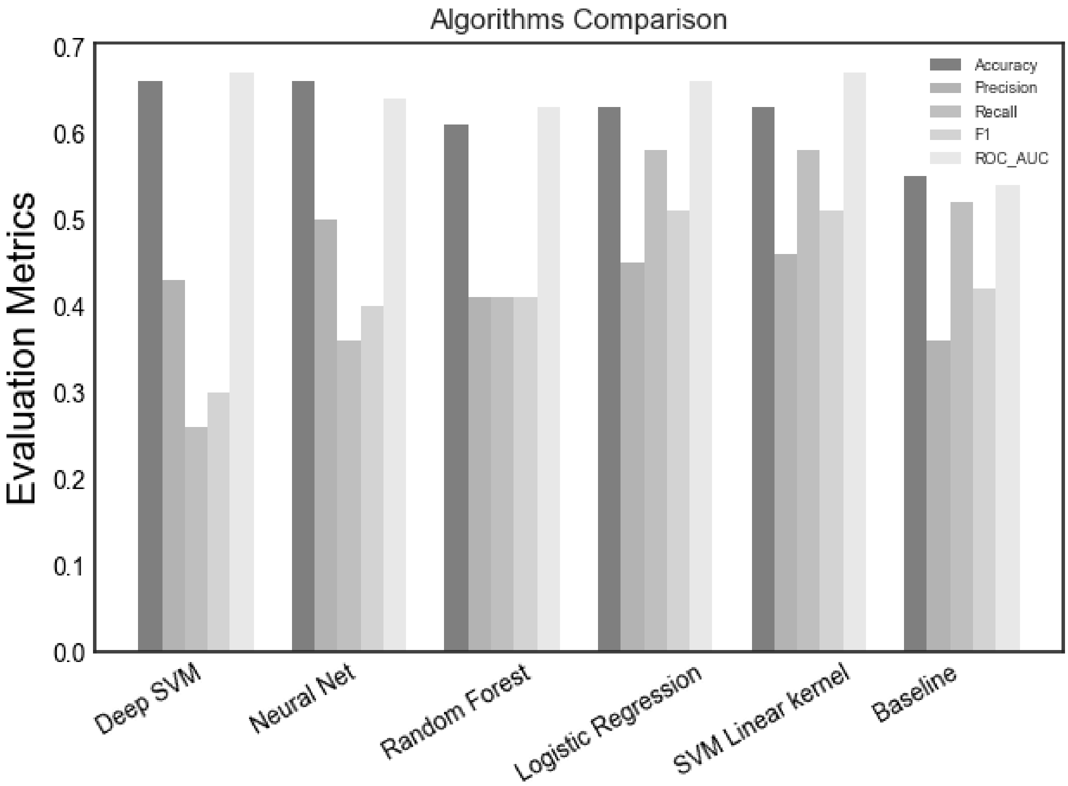

Figure 3: Comparison of different Algorithms for feature set 3

7. Conclusion

Violence experienced during adolescence is related to experiences of violence in adulthood, as a result of developing ineffective coping strategies in situations of conflict within relationships ([6]). Additionally, it is known that children who witness intimate partner violence between their parents are at increased risk of becoming perpetrators themselves. There is a broad understanding that violence is not only learned but there is a cycle of violence meaning violent victimization only perpetuates more violence interpersonally and intergenerationally. To break this cycle of violence, it is imperative that we shift the narrative from victim to perpetrator. Although victim support services are necessary, solely focusing on victim services perpetuates an underlying victim-blaming, misogynistic and oppressive service system. As noted before, perpetrator programs exist but they are ineffective and happen after the incident of violence. Additionally, by switching the narrative from victim to perpetrator we are not proposing a perpetrator-blaming model. To that end, our study aimed to build a triage tool which could help us predict perpetrator behaviors among homeless youth populations before they start showing such behaviors. Our study results indicated that experiencing physical abuse during childhood, being a current or former gang member, and engaging in a recent physical fight were significantly associated with IPV perpetration. Machine learning techniques allowed us to gain better insights into the dataset, which traditional techniques used in social sciences could not have alone.

IPV perpetration is the result of negative early life experiences and other life stressors. Holding the individual accountable does not take into consideration a life-course or systems model which are frameworks that define social work research and practice. To stop the cycle of violence, primary prevention services are necessary. IPV can affect youth regardless of race, culture, or socioeconomic status. However, certain groups are at greater risk of experiencing IPV when compared to the general youth population (OKeefe, 1997) which includes homeless youth. Healthy relationship and conflict resolution skills building programs should be offered widely but in particular within at-risk populations. Given limited resources, which is the unfortunate reality of social service settings, a tool that could triage those that are at-risk for perpetrating violence towards their partner could be extremely useful in reducing overall violence in adolescent and youth populations. This triage tool could be used as brief screening or assessment and provide targeted supportive services that would improve an individual's skills in coping with stress and conflict in a relationship. It is important that predictive models for social good come from a supportive approach rather than punitive. The predictive model and triage tool that we are building will not claim that an individual is a perpetrator it will only identify those that are at-risk. This is an important distinction, especially since our model intends to prevent violent incidents before they occur.

It is important to note some potential limitations of our study. All survey data rely on individuals reports of abuse perpetrated, their experiences of violence may be subject to response bias. Despite extensive interviewee training and efforts to ensure privacy, individuals might not have felt able to disclose their experiences of perpetrating violence to their intimate partners. Although our study provides important insight on what factors are the most important in perpetration of violence, we might have limited details about the degree of violence or the conditions in which the violence occurred, lack of some of which might change our results drastically. Also, the dataset we had was too small to verify the correct applicability of machine learning and data mining algorithms and build a model based on that. In social sciences, a data set with over 300 observations with rich behavioral data is applauded however in computer sciences it is not. This modeling was also only done within a sample of homeless youth in Los Angeles. In the future, we can have more generalized versions of the dataset which spans the demographics, living conditions and age of the individuals. Hopefully, we are able to predict and help the pepetration even better in the coming future with using better datasets and better data mining algorithms.

References

[1] Bandura, A. (1979). The social learning perspective: Mechanisms of aggression.

[2] Delsol, C. and Margolin, G. (2004). The role of family-of-origin violence in men's marital violence perpetration. *Clinical psychology review*, 24(1):99–122.

[3] Eaton, D. K., Kann, L., Kinchen, S., Shanklin, S., Flint, K. H., Hawkins, J., Harris, W. A., Lowry, R., McManus, T., Chyen, D., et al. (2012). Youth risk behavior surveillance-united states, 2011. *Morbidity and mortality weekly report. Surveillance summaries (Washington, DC: 2002)*, 61(4):1–162.

[4] Eriksson, L. and Mazerolle, P. (2015). A cycle of violence? examining family-of-origin violence, attitudes, and intimate partner violence perpetration. *Journal of interpersonal violence*, 30(6):945–964.

[5] Feder, L. and Wilson, D. B. (2005). A meta-analytic review of court-mandated batterer intervention programs: Can courts affect abusers behavior? *Journal of experimental Criminology*, 1(2):239–262.

[6] Flannery, D. J., Singer, M. I., and Wester, K. L. (2003). Violence, coping, and mental health in a community sample of adolescents. *Violence and victims*, 18(4):403–418.

[7] Franklin, C. A. and Kercher, G. A. (2012). The intergenerational transmission of intimate partner violence: Differentiating correlates in a random community sample. *Journal of Family Violence*, 27(3):187–199.

[8] Heerde, J. A., Hemphill, S. A., and Scholes-Balog, K. E. (2014). fightingfor survival: A systematic review of physically violent behavior perpetrated and experienced by homeless young people. *Aggression and violent behavior*, 19(1):50–66.

[9] James, G., Witten, D., and Hastie, T. (2014). An introduction to statistical learning: With applications in r.

[10] Johnson, W. L., Giordano, P. C., Manning, W. D., and Longmore, M. A. (2015). The age–ipv curve: Changes in the perpetration of intimate partner violence during adolescence and young adulthood. *Journal of youth and adolescence*, 44(3):708–726.

[11] Jose, C., Goyal, P., Aggrwal, P., and Varma, M. (2013). Local deep kernel learning for efficient non-linear svm prediction. In *International Conference on Machine Learning*, pages 486–494.

[12] Limbos, M. A., Chan, L. S., Warf, C., Schneir, A., Iverson, E., Shekelle, P., and Kipke, M. D. (2007). Effectiveness of interventions to prevent youth violence: A systematic review. *American journal of preventive medicine*, 33(1):65–74.

[13] Livingston, M. (2011). A longitudinal analysis of alcohol outlet density and domestic violence. *Addiction*, 106(5):919–925.

[14] McGinn, T., Taylor, B., McColgan, M., and Lagdon, S. (2016). Survivor perspectives on ipv perpetrator interventions: a systematic narrative review. *Trauma, Violence, & Abuse*, 17(3):239–255.

[15] McKinney, C. M., Caetano, R., Ramisetty-Mikler, S., and Nelson, S. (2009). Childhood family violence and perpetration and victimization of intimate partner violence: Findings from a national population-based study of couples. *Annals of epidemiology*, 19(1):25–32.

[16] Meyer, S. (2011). Seeking help for intimate partner violence: Victims experiences when approaching the criminal justice system for ipv-related support and protection in an australian jurisdiction. *Feminist Criminology*, 6(4):268–290.

[17] Milburn, N. G., Rotheram-Borus, M. J., Rice, E., Mallet, S., and Rosenthal, D. (2006). Cross-national variations in behavioral profiles among homeless youth. *American journal of community psychology*, 37(1-2):63.

[18] Murphy, C. M. and Baxter, V. A. (1997). Motivating batterers to change in the treatment context. *Journal of Interpersonal Violence*, 12(4):607–619.

[19] Petering, R., Rhoades, H., Rice, E., and Yoshioka-Maxwell, A. (2015). Bidirectional intimate partner violence and drug use among homeless youth. *Journal of interpersonal violence*, page 0886260515593298.

[20] Petering, R., Rice, E., and Rhoades, H. (2016). Violence in the social networks of homeless youths: Implications for network-based prevention programming. *Journal of Adolescent Research*, 31(5):582–605.

[21] Petering, R., Wenzel, S., and Winetrobe, H. (2014). Systematic review of current intimate partner violence prevention programs and applicability to homeless youth. *Journal of the Society for Social Work and Research*, 5(1):107–135.

[22] Press, W. H., Flannery, B. P., Teukolsky, S. A., Vetterling, W. T., et al. (1989). *Numerical recipes*, volume 3. cambridge University Press, cambridge.

[23] Robertson, M. J. and Toro, P. A. (1999). Homeless youth: Research, intervention, and policy. In *Practical lessons: The 1998 national symposium on homelessness research. Washington, DC: US Department of Housing and Urban Development and US Department of Health and Human Services.*

[24] Russell, S. and Norvig, P. (1995). Artificial intelligence: A modern approach. *Artificial Intelligence. Prentice-Hall, Egnlewood Cliffs*, 25:27.

[25] Silverman, J. G., Raj, A., and Clements, K. (2004). Dating violence and associated sexual risk and pregnancy among adolescent girls in the united states. *Pediatrics*, 114(2):e220–e225.

[26] Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(3):273–282.

[27] Tyler, K. A., Melander, L. A., and Noel, H. (2009). Bidirectional partner violence among homeless young adults: Risk factors and outcomes. *Journal of Interpersonal Violence*, 24(6):1014–1035.

[28] Wasserstein, R. L. and Lazar, N. A. (2016). The asa's statement on p-values: context, process, and purpose. *Am Stat*, 70(2):129–133.

[29] Whitbeck, L. B. and Hoyt, D. R. (1999). *Nowhere to grow: Homeless and runaway adolescents and their families.* Transaction Publishers.

[30] Widom, C. S. (1989). The cycle of violence. *Science*, 244(4901):160.

[31] Widom, C. S., Czaja, S., and Dutton, M. A. (2014). Child abuse and neglect and intimate partner violence victimization and perpetration: A prospective investigation. *Child abuse & neglect*, 38(4):650–663.

[32] Widom, C. S. and Wilson, H. W. (2015). Intergenerational transmission of violence. In *Violence and mental health*, pages 27–45. Springer.

[33] Wolfe, D. A. and Jaffe, P. G. (1999). Emerging strategies in the prevention of domestic violence. *The*

future of children, pages 133–144.