Paper ###-2020

School Violence and Safety

Thu Dinh, Alvi Mahmud, Prajakta Pai, Shelby Rangel

Oklahoma State University

ABSTRACT

School violence has become an emerging problem these days in the United States which not only negatively impacts children's educational outcomes but also seriously harms their physical and emotional well-being. School violence also results in wider social-economic costs and long-term impact, since involvement in school violence at smaller ages potentially leads to future antisocial and criminal behaviors, as well as social and relationship difficulties (UNESCO, 2017). Hence, effective action is needed to address the problem of school violence to ensure that all children and adolescents have access to safe and non-violent learning environments. This paper aims to examine the school violence issue in public schools in the United States, to identify the key factors leading to school violence, and to predict the level of violence at a given school. The methods used were to create new features by grouping correlated variables and adjust the bias in the rare target variable's distribution by utilizing oversampling technique before building some classification models within SAS® Enterprise Miner. The model findings indicate that by utilizing the logistic regression model, approximately 20.8% misclassification rate, 79.3% sensitivity rate, and 78.7% specificity rate at distinguishing whether the school has violent incident were achieved. The model can be further generalized to apply on higher education institutions like colleges and universities. The paper's findings are expected to benefit parents in selecting a safe school for their children and also to facilitate researchers, school authorities and policymakers in evaluating the effectiveness of current school policies and practices on violence issues and assisting them in revising strategies to prevent violence and create a safe educational environment.

INTRODUCTION

BACKGROUND

School violence is defined as physical violence, psychological violence, sexual violence, and bullying (UNESCO, 2017). Many such problems arise from antisocial behavior, family upbringing and even access to fire arms along with peer pressure. Numerous incidences of cyberbullying and psychological torture have been recorded over the years and have been observed to be increasing in number (UNESCO, 2017). Almost 9% of students had been in one or more physical fights on school property while nearly 6% had been threatened on school property in 2016-2017 (CDC, 2017). In 46 weeks of 2019, there have been 45 shootings at US schools in which someone was hurt or killed (Wolfe and Walker, 2019). All of these events are considered school violence which affects the students attending the school.

DATA DESCRIPTION

The School Survey on Crime and Safety (SSOCS) is the primary source of school-level data on crime and safety; it is managed by the National Center for Education Statistics (NCES) within the Institute of Education Sciences of the U.S. Department of Education and is available for download at https://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2018109. The SSOCS is a nationally representative cross-sectional survey of about 4,800 public elementary and secondary schools which has been conducted six times within 1999-2016.

The dataset chosen for this analysis only covers the latest administration of SSOCS 2015-2016, in which data were collected from February 22, 2016 through July 5, 2016. The dataset has 2,092 rows (the number of public schools that submitted completed questionnaires) with a total of 476 variables, including 230 unique variables, 50 Jackknife replicate variables (that produce national estimates from the listed variables), and 196 imputation flags (explaining if and how imputations were performed by the NCES). In term of variable types, there are 101 binary, 89 interval, 203 nominal, and 83 ordinal variables. The data covers such topic as school practices and programs, school security staff, school mental health services, parent/community involvement, limitations on crime prevention, disciplinary problems and actions, and other school characteristics related to school crime.

PROBLEM STATEMENT

Violence and bullying in schools violate the laws and regulations put in place by the schools to help give students a safe educational environment. Not only does school violence have a negative impact on the physical and mental health, and emotional well-being of those who are victimized, it also has a detrimental effect on the wider school environment (Ferrara et al., 2019). More seriously, school violence can also lead to wider social-economic costs and long-term impact as the effects persist into adult life. In light of these potentially negative effects, this study aims at exploring the most important factors that leads to school violence and predicting the level of violence in a given school. The paper's findings can help parents make better informed decisions when enrolling their kids in school by learning first hand from the schools how they prevent and control school violence. The School Survey on Crime and Safety (SSOCS) is a school-level survey where a school representative has an opportunity to answer questions about their safety procedures as well as crime incidents that happen within their school. In addition, understanding school crimes and contributing factors will also help researchers, school authorities and policymakers identify trends in crime and safety issues. This will facilitate the process of addressing school violence issues and devising strategies to prevent and tackle them in order to create a secure and effective educational environment.

METHOD

DATA CLEANING/ VALIDATION

Since the focus of the study is predicting the level of violence in a given school, the VIOINC16 variable, which is the *total number of school violent incidents reported*, was selected as the target variable. Due to dimensionality and limited scope of the study, all the Jackknife replicate variables and the imputation flags variables were not used in analysis. In the remaining 230 variables, there exist some variables that are highly correlated with the target variable, for example: INCID16 (Total number of incidents recorded), SVINC16 (Total number of serious violent incidents recorded) and VIOPOL16 (Total number of violent incidents reported to police). These variables were dropped. The dataset was then partitioned into 80% for training and 20% for validation.

FEATURE ENGINEERING

The interval target variable VIOINC16 was converted to a binary target (named VIOINC) where VIOINC = 0 when the original target VIOINC16 has zero (0) value, and VIOINC = 1 when the original target VIOINC16 has a greater-than-zero value. Most of the input variables which are either binary or ordinal were grouped or combined into new features given their correlations and meanings. The list of all 34 newly created features is provided in Appendix A. After data cleaning and feature engineering, the total number of variables was reduced from 476 to 90 variables.

OVERSAMPLING

Since the distribution of the target variable (VIOINC16) was imbalanced with 17.35% of the schools having no violent incidents, oversampling was utilized to make the proportion of events and non-events more balanced. After splitting the data into a training and a validation set, the minority class (VIOINC = 0) was oversampled using PROC SURVEYSELECT with unrestricted random sampling method. The oversampled data were appended to the existing training data to create new training data with approximately equal proportions in the target class.

RESULTS

EXPLORATORY ANALYSIS

Exploratory analysis of some demographic variables was conducted to examine if there were noticeable differences across schools. As an example, a visualization of the distribution of two violence levels (VIOINC = 0 and VIOINC = 1) across different levels of school size and school type is provided in Figure 1.

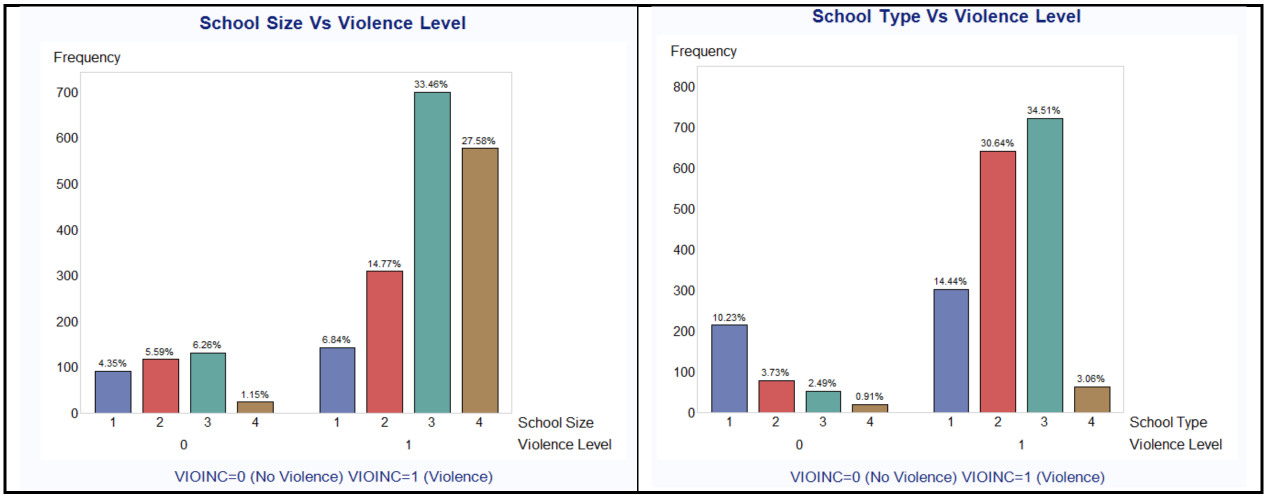


Figure 1. Comparison in size and type between violent schools and non-violent schools

Figure 1 indicates that smaller schools tend to be less violent while the bigger schools tend to be more violent. Also, middle and high schools (School type = 2 and 3) seem to have more violent incidents than primary schools (School type = 1) or combined schools (School type =4). Similarly, number of transfers in and out of the school, location of the school, and the frequency of cyberbullying occurrences were also found to have different distributions for violent and non-violent schools.

MODELS

The Weight of Evidence method was used to reduce the number of variables using the Interactive Grouping node inside SAS Enterprise Miner. The model misclassification rate improved about 2% with that. With Information Value cutoff set to 0.1, the number of variables was reduced to 37. Several machine learning models in SAS Enterprise Miner including Random Forest, Logistic Regression and Gradient Boosting were employed. The purpose of choosing these models was to identify predictive power of variables for the target and all three models provided such statistics in the form of variable importance or odds ratio. The results of the models for validation data were reported in Table 1 below:

Model	Misclassification	Sensitivity	Specificity
Logistic Regression	20.76%	79.33%	78.69%
Random Forest	19.33%	82.40%	70.49%
Gradient Boosting	23.39%	75.70%	81.97%

Table 1. Classification models results

Although Random Forest produced better accuracy, it had poor specificity compared to other models, thus there was one possibility that the Random Forest algorithm was over fitting. Gradient Boosting did not perform well as expected, possibly because it is more data hungry than other algorithms and there is limited variation in the data (Van, T., 2014). Logistic

Regression had a good accuracy, sensitivity and specificity overall (refer to Appendix B for model classification results), hence it was selected as the best fit model for the dataset. The significant variables from the Logistic Regression model (refer to Appendix C for a visualization of variable importance) include Cyberbullying, FR_SIZE (School size), FR_LVEL (School type), C0536 (percentage of students who consider academic achievements important), school policies and procedures (Inside_Suspend_With_Services, Outside_Suspend_No_Services, Outside_Suspend_With_Services), Community_Involvement, Disruption_level (level of perceived disruption), and Transfer_Specialize (Transfer to specialized schools for special violence allowed and/or used). Most of these variables were also found to be significant in the Random Forest and Gradient Boosting model.

The Odds Ratio estimates from Logistic Regression Model (Appendix D) indicate that larger schools with more than 1000+ students (FR_SIZE =4) have higher odds of being violent. As the level of cyberbullying incidents increases (the students and school are affected with cyberbullying more frequently), the odds of the school being violent tend to increase. It can also be seen that as the suspension policies and transfer policies increases (policy exists and used), the odds of the school being violent decreases. However, it is not clear whether those policies reduce violence or the school violence paves way for those policies. The odds ratio suggests that if more students take academic achievements importantly (i.e., variable C0536 increases), the odds of the school being a violent one decrease. Similarly, as Disruption level (which is a variable that indexes the perceived level of disruption in school due to sexual crime, race related tension among other violent events) is more frequent, the odds of the school being violent increases.

GENERALIZATION

The takeaways from the research can aid parents and government bodies in charge of educational reforms alike. There are lots of tools parents could use to find out well performing schools and past history can help to determine whether the school environment is violent or not. This research offers a different approach as it uses survey data from teachers instead of students/parents which is more common. For example, a key takeaway from the exploratory data analysis is high number of transfers in and out for violent schools is something that the popular rankings do not tend to embrace (Morse, R., 2019; College Stats, 2012), which can help parents determine whether a school is non-violent or not. There is also extensive research that shows the impact of student to teacher ratio (Rotham, R., 2003) and small school size on school performance (Hylden, J., 2005). The result from the research further corroborates that smaller schools and less frequent change of classrooms are characteristics of less violent schools (number of teachers was not present in the dataset though), and officials in charge of school districts can use this information to find the optimum size of schools and classrooms to tackle school violence.

FUTURE STUDIES

With more data covering longer time spans, a time series analysis could be performed, which will allow researchers to see how the policies undertaken by schools throughout the country have changed and what impact they have on violence. If school size or demographics remained unchanged, this would allow future researchers to be more precise about the causality of policies and actions. Additionally, future research can be conducted to predict transfers of students and examine the effects leading to the transfers to tackle that. There have been some studies analyzing the causes of the transfers (Rumberger, R., 2003) but no significant research has been done on the modeling aspect.

CONCLUSION

The paper’s analysis and findings highlight the key factors that have significant impact on the number of violent incidents a school endures; these factors include: the number of students that attend the school, the grade level of the school, the level of cyberbullying incidents, the percent of students who find academic achievement important, community involvement, and the school's practices and policies on school violence. Other factors, such as the crime level in each school district and the student demographics, have minimal impact on future violent incident predictions. Given these findings, a couple of recommendations were suggested as below:

- Reduce the school size (the number of students who attend the school) and decrease the frequency of classroom changes throughout the day
- Create awareness about cyberbullying and its effects on students within the teachers and parents and include cyberbullying as an indicator for school violence
- Promote greater participation from the community in school events and activities
- Implement policies and procedures that have been proven to decrease the number of violent incidents in schools that have not yet been implemented

These insights and recommendations could help reduce the violence within schools.

REFERENCES

CDC. 2017. "Preventing School Violence". Accessed December 4, 2019. https://www.cdc.gov/violenceprevention/youthviolence/schoolviolence/fastfact.html.

Ferrara, P., Franceschini, G., Villani, A., Corsello, G. 2019. "Physical, psychological and social impact of school violence on children". *Italian Journal of Pediatrics*. 45 (76).

Hylden, J. 2005. "What's So Big about Small Schools? The Case for Small Schools: Nationwide and in North Dakota". *Program on Education Policy and Governance*. PEPG 05-05, 51 pp.

Jackson, M., Diliberti, M., Kemp, J., Hummel, S., Cox, C., Gbondo-Tugbawa, K., Simon, D., and Hansen, R. 2018. *2015–16 School Survey on Crime and Safety (SSOCS).* U.S. Department of Education, National Center for Education Statistics. Washington, DC.

Moose, R., Brooks, E. 2019. "How U.S. News Calculated the 2019 Best High Schools Rankings". Accessed November 15, 2019. https://www.usnews.com/education/best-high-schools/articles/how-us-news-calculated-the-rankings

Rotham, R. 2003. "Transforming high schools into small learning communities". *Challenge Journal*. 6(1), 1-8.

Rumberger, R. 2003. The Causes and Consequences of Student Mobility. *The Journal of Negro Education, 72*(1), 6-21. doi:10.2307/3211287

UNESCO. 2017. *School Violence and Bullying: Global Status Report*. ISBN 978-92-3-100197-0.

Van, T., Austin, P.C. & Steyerberg, E.W. 2014. "Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints". *BMC Med Res Methodol* 14, 137. doi:10.1186/1471-2288-14-137.

Wolfe, E., Walker, C. "In 46 weeks this year, there have been 45 school shootings". *CNN*. Accessed October 21, 2019. https://www.cnn.com/2019/05/08/us/school-shootings-us-2019-trnd/index.html

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Thu Dinh
Oklahoma State University
thu.dinh@okstate.edu

Alvi Mahmud
Oklahoma State University
alvi.mahmud@okstate.edu

Prajakta Pai
Oklahoma State University
prajakta.pai@okstate.edu

Shelby Rangel
Oklahoma State University
shelby.rangel@okstate.edu

APPENDIX A: LIST OF NEW VARIABLES CREATED FROM GROUPING

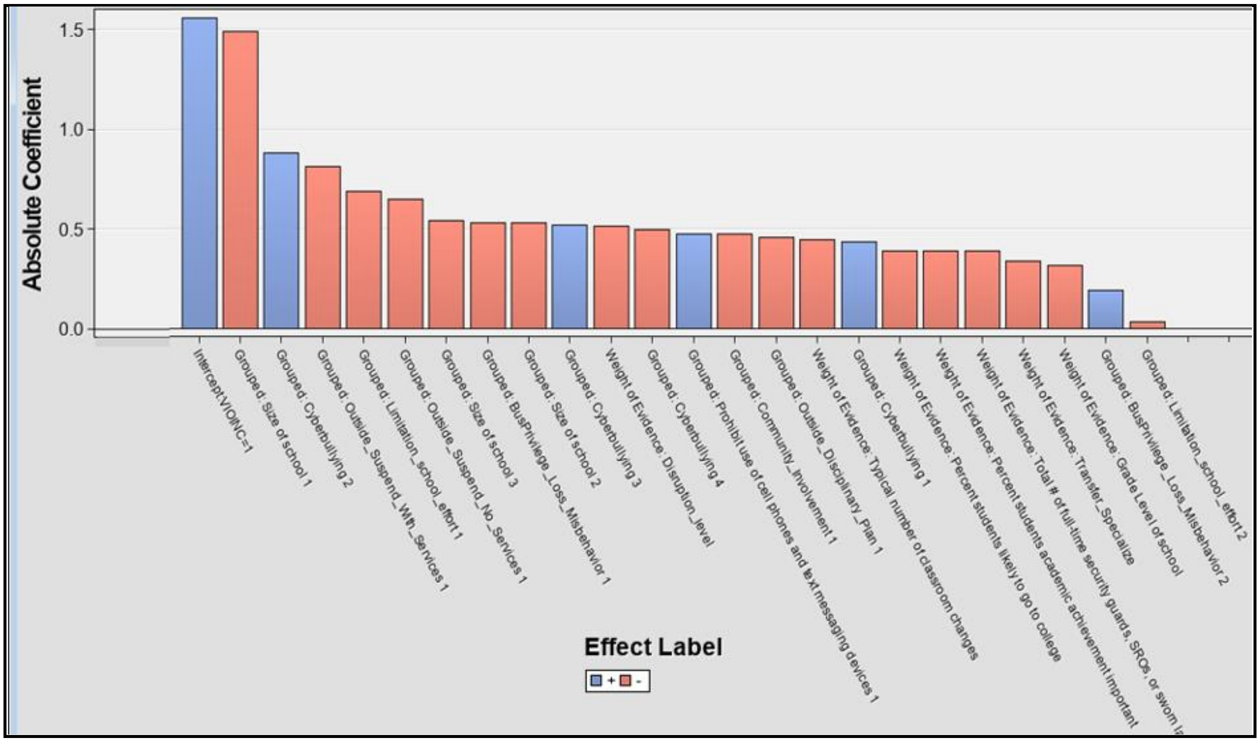
New variable	Type	Grouped from original variables
Access_Controlled	Ordinal	C0112, C0114
Drug_Testing	Ordinal	C0128, C0130
Written_Plans	Ordinal	C0155, C0157, C0158, C0162, C0166, C0169, C0170, C0173
Drilled_Plans	Ordinal	C0163, C0165, C0167
Acceptance_Groups	Ordinal	C0604, C0606, C0608
Community_Involvement	Interval	C0204-C0218
Law_Enforcement_Presence	Interval	C0610, C0612-C0618, C0648
Law_Enforcement_Equipped	Interval	C0610, C0620-C0626
Law_Enforcement_Participation	Interval	C0610, C0628-C0646
Mental_Health_Assessment	Interval	C0662-C0666
Mental_Health_Treatment	Interval	C0668-C0672
Mental_Health_Efforts_Limits	Interval	C0674-C0686
Teacher_Training	Interval	C0265-C0277
Removal_No_Services	Ordinal	C0390, C0392
Removal_Tutoring	Ordinal	C0394, C0396
Transfer_Specialize	Ordinal	C0398, C0400
Transfer_Regular	Ordinal	C0402, C0404
Outside_Suspend_No_Services	Ordinal	C0406, C0408
Outside_Suspend_With_Services	Ordinal	C0410, C0412
Inside_Suspend_No_Services	Ordinal	C0414, C0416
Inside_Suspend_With_Services	Ordinal	C0418, C0420
Referral_School_Counselor	Ordinal	C0422, C0424
Inside_Disciplinary_Plan	Ordinal	C0426, C0428
Outside_Disciplinary_Plan	Ordinal	C0430, C0432
BusPrivilege_Loss_Misbehavior	Ordinal	C0434, C0436
Corporal_Punishment	Ordinal	C0438, C0440
School_Probation	Ordinal	C0442, C0444
Detention_Saturday_School	Ordinal	C0446, C0448
Student_Privileges_Loss	Ordinal	C0450, C0452
Require_Community_Service	Ordinal	C0454, C0456
Disruption_level	Interval	C0374-C0386
Cyberbullying	Interval	C0389-C0393
Limitation_school_effort	Interval	C0280-C0304
Parent_participation	Interval	C0196-C0202

APPENDIX B: LOGISTIC REGRESSION MODEL CLASSIFICATION TABLE

Data Role=VALIDATE Target Variable=VIOINC Target Label=' '

Target	Outcome	Target Percentage	Outcome Percentage	Frequency Count	Total Percentage
0	0	39.3443	78.6885	48	11.4558
1	0	60.6557	20.6704	74	17.6611
0	1	4.3771	21.3115	13	3.1026
1	1	95.6229	79.3296	284	67.7804

APPENDIX C: VARIABLE IMPORTANCE FROM LOGISTIC REGRESSION



APPENDIX D: ODDS RATIO ESTIMATES FROM LOGISTIC REGRESSION

Effect		Point Estimate
Cyberbullying	1 vs 5	2.364
Cyberbullying	2 vs 5	1.983
Cyberbullying	3 vs 5	1.400
Cyberbullying	4 vs 5	0.624
FR_SIZE	1 vs 4	0.266
FR_SIZE	2 vs 4	0.616
FR_SIZE	3 vs 4	0.673
Inside_Suspend_With_Services	1 vs 2	0.595
Outside_Suspend_No_Services	1 vs 2	0.538
Outside_Suspend_With_Service	1 vs 2	0.505
C0536		0.592
Community_Involvement		0.563
Disruption_level		0.574
FR_LVEL		0.671
Transfer_Specialize		0.608